Rapport

Opgave AI

De nieuwe systeemtechnologie

WRR

Opgave AI
De nieuwe systeemtechnologie

Opgave AI. De nieuwe systeemtechnologie is een advies aan de regering uit naam van de voltallige Wetenschappelijke Raad voor het Regeringsbeleid. WRR-Rapport 105 is voorbereid en geschreven door:

Prof. mr. J.E.J. (Corien) Prins (raadsvoorzitter),
Prof. dr. H. (Haroon) Sheikh (projectcoördinator),
Dr. E.K. (Erik) Schrijvers (projectcoördinator),
Drs. E.L. (Eline) de Jong (projectmedewerker),
Mr. J.M. (Monique) Steijns (projectmedewerker),
Prof. dr. mr. M.A.P. (Mark) Bovens (raadslid).

De Wetenschappelijke Raad voor het Regeringsbeleid werd in voorlopige vorm ingesteld in 1972. Zijn positie is definitief vastgelegd bij wet van 30 juni 1976 (Stb. 413). De Wetenschappelijke Raad voor het Regeringsbeleid (WRR) is een onafhankelijk adviesorgaan. De WRR informeert en adviseert de regering en het parlement over sectoroverstijgende vraagstukken die grote impact hebben op de samenleving. De adviezen zijn gebaseerd op wetenschappelijke onderzoek en gericht op een langetermijnperspectief.

De huidige zittingsperiode loopt tot 31 december 2022. De samenstelling van de raad is:
Prof. dr. mr. C .C .J.H. (Catrien) Bijleveld,
Prof. dr. A .W. A . (Arnoud) Boot,
Prof. dr. mr. M. A .P. (Mark) Bovens,
Prof. dr. G.B.M. (Godfried) Engbersen,
Prof. dr. S.J.M.H. (Suzanne) Hulscher,
Prof. mr. J.E .J. (Corien) Prins (voorzitter),
Prof. dr. M. (Marianne) de Visser,
Prof. dr. C .G. (Casper) de Vries,
Secretaris: Prof. dr. F.W.A. (Frans) Brom.

Rapport

Opgave AI

De nieuwe systeemtechnologie

WRR

WETENSCHAPPELIJKE RAAD VOOR HET REGERINGSBELEID

Redactie: Simone Langeweg, Leiderdorp
Uitgever: WRR

Vormgeving binnenwerk: VormVijf, Den Haag
Omslagafbeelding: Steffie Padmos, Amsterdam
Figuren en tabellen: VormVijf, Den Haag

ISBN 978-90-832012-3-8
NUR 984

Wetenschappelijke Raad voor het Regeringsbeleid
Buitenhof 34
Postbus 20004
2500 EA Den Haag
info@wrr.nl
wrr.nl

WRR

WETENSCHAPPELIJKE RAAD VOOR HET REGERINGSBELEID

Aan de Minister-President
Voorzitter van de Ministerraad
De heer drs. M. Rutte
Postbus 20001
2500 EA Den Haag

ons kenmerk
2021/CP/FB

telefoonnummer
070 356 4694

onderwerp
WRR-rapport nr. 105
Opgave AI. De nieuwe systeemtechnologie

e-mail
secretariaat@wrr.nl

datum
5 november 2021

Het doet ons genoegen u hierbij het rapport *Opgave AI. De nieuwe systeemtechnologie* aan te bieden. Met dit rapport geeft de WRR antwoord op de adviesaanvraag van het kabinet over het thema 'AI en publieke waarden'.

De WRR concludeert dat de overheid zich actiever moet voorbereiden op een samenleving waarin Artificiële Intelligentie (AI) een grote rol speelt. Vijf door de WRR geïdentificeerde opgaven dienen daarbij richtinggevend te zijn. Deze opgaven vloeien voort uit het bijzondere karakter van AI. AI is niet zomaar een technologie. Zij valt het beste te vergelijken met de opkomst van de stoommachine, elektriciteit, de verbrandingsmotor en de computer. Dit zijn systeemtechnologieën, die door de hele economie en samenleving voor allerlei doelen gebruikt kunnen worden. Dergelijke technologieën zijn alomtegenwoordig en hebben een grote en onvoorspelbare impact. Welke publieke waarden door AI worden geraakt en op welke wijze, valt onmogelijk vooraf te bepalen. Wel is duidelijk dat de rol van de overheid zal groeien naarmate AI verder ingebed raakt in de samenleving. De vijf opgaven die de WRR destilleert uit de omgang met eerdere systeemtechnologieën en waarvoor de overheid staat bij de inbedding van AI in de Nederlandse samenleving zijn: 1) Demystificatie: het adresseren van onrealistische beelden; 2) Contextualisering: het creëren van goede omgeving om AI te laten functioneren; 3) Engagement: het betrekken van maatschappelijke partijen bij de technologie; 4) Regulering: het opstellen van kaders en 5) Positionering: het strategisch nadenken over de verhouding van Nederland ten opzichte van partijen elders.

De overheid dient zich nu al voor te bereiden op deze vijf opgaven. Zo moet de overheid leren AI tot een vast onderdeel van het eigen functioneren te maken en inzage te geven in algoritmegebruik via algoritmeregisters. Om de contextualisering van AI meer richting te geven beveelt de WRR een Nederlandse 'AI-identiteit' aan en een stelsel van opleiding en certificering. Daarnaast is het nodig de capaciteit van belangenorganisaties te versterken, zodat zij mede sturing kunnen geven aan maatschappelijk wenselijk AI-gebruik. Een goede terugkoppeling tussen de ontwikkelaar van AI, de gebruiker ervan en de personen die er in de praktijk de consequenties van ondervinden, is daarbij onmisbaar. Kaders voor AI moeten zich behalve op de technologie zelf ook richten op factoren die de inbedding ervan beïnvloeden, zoals grootschalige dataverzameling en monopolies. Om deze thema's integraal te kunnen adresseren, beveelt de WRR een brede strategische wetgevingsagenda aan. Tot slot zal de overheid op internationaal vlak moeten investeren in AI-diplomatie en in meer kennis omtrent de veiligheidsrisico's van AI.

Het kabinet heeft de WRR tot slot ook de vraag voorgelegd of het huidige instrumentarium voor AI voldoende adequaat is. In antwoord hierop bepleit de WRR dat de overheid een andere, meer solide beleidsinfrastructuur ontwikkelt, te beginnen met een AI-coördinatiecentrum. De WRR acht het van wezenlijk belang dat deze infrastructuur behalve ambtelijk vooral ook politiek is verankerd. Met deze stappen kan de Nederlandse overheid bijdragen aan de inbedding van AI, de verbrandingsmotor van de 21e eeuw.

Ingevolge de Instellingswet ziet de raad graag de bevindingen van de ministerraad tegemoet.

De voorzitter,

Prof. mr. J.E.J. (Corien) Prins

De secretaris,

Prof. dr. F.W.A. (Frans) Brom

Buitenhof 34, Postbus 20004, 2500 EA Den Haag
telefoon (070) 356 46 00, website: www.wrr.nl, e-mail: info@wrr.nl

Inhoud

Deel 2

Vijf opgaven: bespreking van de opgaven voor de maatschappelijke inbedding van AI

Deel 3
Agenda: conclusies en aanbevelingen voor AI-beleid in Nederland

Samenvatting

AI is niet zomaar een technologie, maar een systeemtechnologie die de samenleving fundamenteel zal veranderen. Dat is de hoofdboodschap van dit rapport. Een systeemtechnologie is alomtegenwoordig, kent continue verbetering en maakt complementaire innovatie mogelijk. De ontwikkeling van deze technologie staat momenteel op een keerpunt: de overgang van het lab naar de samenleving, waarin de technologie met de tijd ingebed moet raken. Uit de geschiedenis van eerdere systeemtechnologieën valt te leren dat dat proces van inbedding uiteenvalt in vijf opgaven voor overheid en samenleving. Alleen wanneer de overheid deze opgaven structureel ter hand neemt, kan zij de publieke waarden die gemoeid zijn met de inzet van AI ook in de toekomst beschermen en garanderen. Dat vergt een beleidsinfrastructuur waarin zowel de politieke als de ambtelijke inzet verankerd zijn.

AI op een keerpunt

We staan wat betreft AI momenteel op een keerpunt. Het begin van AI als discipline dateert van 1956, toen wetenschappers het Dartmouth Summer Research Project on Artificial Intelligence organiseerden. In het lab heeft de technologie vervolgens verschillende golven doorgemaakt, waarvan de recentste – beginnend in de jaren negentig – op dit moment in een stroomversnelling geraakt. Wetenschappelijke doorbraken binnen het domein van *machine learning*, en daarbinnen *deep learning*, vormen samen met de toename van rekenkracht en beschikbare data, de stuwende kracht achter deze golf. Na 2010 maakt AI definitief en breed de overstap van de wetenschappelijke wereld van het lab naar de praktijk van onze samenleving. Het aantal AI-patenten stijgt, investeringen in AI nemen toe en ook overheden zetten erop in, waardoor er niet alleen in het bedrijfsleven maar ook bij de overheid meer en meer toepassingen van de technologie te vinden zijn.

Met de overstap van het lab naar de samenleving, evolueert AI ook als maatschappelijk fenomeen. De brede toepasbaarheid maakt de technologie interessant voor partijen in tal van maatschappelijke sectoren. Uit nationale AI-strategieën blijkt dat overheden in AI een belangrijke bron zien van toekomstige economische groei, manieren om de eigen dienstverlening te verbeteren, maar ook een fenomeen met risico's, waarvoor regulering en toezicht ontwikkeld dient te worden. Breder in de samenleving barst een discussie los over allerlei toekomstbeelden en ontstaan controverses rond het gebruik van AI. Het belang en de dynamiek van het lab nemen daarbij overigens geenszins af. Fundamenteel onderzoek blijft nodig, zowel om allerlei beperkingen en tekortkomingen te overwinnen als om de technieken door te ontwikkelen. Kortom: ook het lab hoort nog steeds bij AI.

AI is een systeemtechnologie

Bijzonder aan AI is haar brede toepasbaarheid. De WRR hanteert in dit rapport voor AI de definitie van de Europese High-Level Expert Group on AI: "systemen die intelligent gedrag vertonen door hun omgeving te analyseren en – met enige graad van autonomie – actie te ondernemen om specifieke doelen te bereiken." Vanwege het breed toepasbare karakter van AI typeert de WRR AI als een 'systeemtechnologie'. AI is in onze eeuw wat elektriciteit was in de negentiende eeuw, of de verbrandingsmotor in de twintigste eeuw. Ze is geen concrete technologie die goed te overzien is en door een groep van experts of beleidsmedewerkers van een of enkele ministeries in goede banen valt te leiden. Omdat AI alomtegenwoordig is, continue verbetering kent en complementaire innovatie mogelijk maakt, is ze een veelzijdig en deels onvoorspelbaar fenomeen.

De inbedding van een systeemtechnologie als AI vergt een langdurige wisselwerking van samenleving en technologie. Welke publieke waarden erdoor worden geraakt en op welke wijze, valt daarbij onmogelijk bij voorbaat af te bakenen en is bovendien niet eenduidig. AI zal invloed hebben op onder andere veiligheid en gezondheid, autonomie en vrijheid, burgerrechten en de rechtstaat, rechtvaardigheid en inclusie. Dat zal echter op manieren gebeuren die we nu grotendeels nog niet kunnen voorzien.

De WRR betoogt dat de inbedding van een systeemtechnologie als AI vijf opgaven met zich meebrengt voor overheid en samenleving. Deze opgaven zijn: 1) *demystificatie* van wat AI is en kan; 2) *contextualisering* van de ontwikkeling en toepassing van AI; 3) *engagement* van verschillende partijen; 4) *regulering* van technologie en data, het gebruik ervan en maatschappelijke implicaties; en 5) *positionering* ten opzichte van andere landen en internationale organisaties. Bij elk van deze opgaven bepleit de WRR in dit rapport een transitie en koppelt daaraan, per opgave, twee aanbevelingen.

Demystificatie: Van beeld naar begrip

De opgave 'demystificatie' betreft de beelden die er over AI als technologie bestaan. Centraal staat hier de vraag: *Waar hebben we het over?* Rondom systeemtechnologieën ontstaan altijd extreme beelden. Bij AI is dat de gedachte dat AI-systemen rationeel en objectief zijn en werken als een volstrekt onbegrijpelijke '*black box*'. Ook leeft het idee dat de technologie alle menselijke vermogens zou kunnen evenaren en zelfs overstijgen. AI zou zich zelfs tegen de mensheid kunnen keren. Eerder al bestonden er mythen over bredere digitalisering, zoals dat de ontwikkeling van het internet vrij moet worden gelaten, dat er geen alternatieven zijn voor de huidige vorm van digitale technologie en dat digitalisering een oplossing biedt voor vrijwel ieder probleem.

Te hooggespannen verwachtingen leiden echter tot desillusie en ondoordachte toepassingen, terwijl overtrokken angsten leiden tot afkeer van de technologie en het niet benutten van de kansen die ze biedt. Vooral op de langere termijn zal het vasthouden aan dergelijke beelden negatief uitwerken. De WRR stelt dat meer realisme nodig is om maatschappelijk, en vooral ook met het oog op publieke waarden, de juiste vragen te kunnen stellen: een transitie is nodig van beelden over AI naar begrip van AI.

Om deze transitie te stimuleren, moet de overheid leren over AI tot integraal onderdeel van haar functioneren maken. Zij moet zich bovendien kritisch opstellen wanneer partijen met hoge verwachtingen over de mogelijkheden van AI spreken of hierover risicovolle scenario's schetsen. Dit vergt om te beginnen meer nadruk op het aantrekken van talent en de scholing van personeel. Daarnaast is meer zorg vereist voor basale zaken als tijdige en zorgvuldige archivering bij processen waar AI wordt ingezet, overdracht van kennis als mensen van baan veranderen en toegang tot databases en algoritmen. Ten slotte zijn er implicaties voor de werkwijze van de overheid: een meer iteratief proces, met kleinere projecten, valt te prefereren boven grote IT-projecten met een vaste opleverdatum. Overigens zijn niet alleen de uitvoerende instanties binnen de overheid gebaat bij een lerende aanpak gebaseerd op voortschrijdend inzicht met AI, dat geldt evenzeer voor de politiek, de wetgever alsmede toezichthoudende en rechtsprekende instanties.

Aanbeveling 1
Maak leren over AI en de toepassing daarvan tot een expliciet doel bij het handelen door de overheid.

Ook de bredere samenleving is gebaat bij demystificatie en dus bij een realistischer begrip van AI, oftewel 'AI-wijsheid'. Een tweede stap die de overheid daarom zou moeten zetten, is meer openheid te geven over het eigen gebruik van AI, en wel door initiatieven met algoritmeregisters op te zetten en verder uit te bouwen. De instelling van algoritmeregisters heeft echter pas echt meerwaarde als de overheid daarmee ook het gesprek over AI-gebruik in gang zet – zowel met degenen die de toepassingen gebruiken als met degenen die ermee te maken krijgen. Of dit daadwerkelijk gebeurt, hangt sterk af van de kwaliteit van de geboden informatie, het vermogen van de samenleving om hiermee aan de slag te gaan en de reactie van de verantwoordelijke partijen op de geconstateerde problemen. De instelling van algoritmeregisters moet daarom gepaard gaan met de verplichting ze periodiek te evalueren.

Aanbeveling 2
Stimuleer als overheid de ontwikkeling van AI-wijsheid bij het brede publiek, te beginnen met het opzetten van algoritmeregisters.

Contextualisering: Van techniek naar toepassing

De opgave van contextualisering betreft de toepassing van AI en gaat over de vraag: *Hoe gaat het werken?* Contextualisering impliceert aandacht voor de verbinding van AI met het bredere technische ecosysteem. AI vereist verschillende ondersteunende technologieën, zoals telecommunicatienetwerken, chips en supercomputers. Ook vraagt AI om kwalitatief goede data en rechtszekerheid over geoorloofd gebruik daarvan. Bovendien raakt AI verbonden met andere nieuwe technologieën, zoals 5G-netwerken, het Internet of Things en *quantum computing*. Contextualisering impliceert daarnaast een proces van intensieve training en oefening om AI op de werkvloer effectief te kunnen maken, met als belangrijke vraag hoe een goede mens-machine-interactie tot stand kan komen. Te weinig aandacht voor ondersteunende technologieën en data leidt ertoe dat AI-systemen slecht zullen functioneren, en dat kansen onvoldoende worden benut dan wel verdere ontwikkeling stagneert. Verwaarlozing van het sociale ecosysteem resulteert in gebrekkige implementatie en misstanden of zelfs afwijzing van de technologie omdat de gebruikers van AI-systemen onvoldoende geëquipeerd zijn. De WRR bepleit daarom om niet alleen de techniek zelf maar ook de toepassing ervan centraal te stellen in het beleid.

Het is voor de overheid ondoenlijk om de helpende hand te bieden in alle domeinen waar AI-toepassingen hun weg zouden kunnen vinden. Bovendien kunnen bedrijven veel aanpassingen ook zelf uitstekend doorvoeren. Toch is het risico reëel dat in sommige voor ons land belangrijke economische sectoren de noodzakelijke aanpassingen van de context waarin AI tot ontwikkeling moet komen, niet of onvoldoende ter hand worden genomen. Ook is voorstelbaar dat AI-ontwikkelingen niet goed van de grond komen in publieke sectoren, waar ze juist van groot maatschappelijk belang zijn. De WRR pleit daarom voor de ontwikkeling van een Nederlandse 'AI-identiteit' die uit de domeinen bestaat waarop ons land AI wil ontwikkelen en inzetten. De overheid kan de Nederlandse AI-identiteit ook ondersteunen door strategisch gebruik te maken van het aanbestedingsbeleid.

Aanbeveling 3
Kies expliciet voor een Nederlandse AI-identiteit en onderzoek waar in de betreffende domeinen aanpassingen aan de technische omgeving nodig zijn.

Vervolgens dient in de praktijk meer aandacht uit te gaan naar mens-machine-interactie. De verschillende AI-labs die momenteel gestalte krijgen, kunnen hier een belangrijke rol bij spelen. De overheid kan ook in het eigen gebruik van AI sturen op meer aandacht voor de dynamiek tussen mens en machine. Daarnaast dienen de gedragscontext en de eisen die gelden voor het gebruik van AI, een vast onderdeel te worden van interne toezichtprocessen en richtlijnen. Een goede mens-machine-interactie verlangt behalve certificering op het niveau van een product of organisatie echter ook certificering op het niveau van het individu. De WRR beveelt aan om hiervoor een stelsel van opleiding en certificering te ontwikkelen. Overigens dient niet iedereen die met AI te maken krijgt, geschoold te zijn en over een vaardigheidsbewijs of AI-brevet te beschikken, maar wel degenen die deze technologie hanteren dan wel voor de inzet daarvan verantwoordelijk zijn. Van hen mag worden verlangd dat ze de daartoe benodigde kennis en vaardigheden kunnen aantonen.

Aanbeveling 4
Versterk de vaardigheden en het kritische vermogen van individuen die met AI-systemen werken en ontwikkel daarvoor een stelsel van opleiding en certificering.

Engagement: Van monoloog naar dialoog

De opgave van engagement betreft de maatschappelijke omgeving van AI en gaat over de vraag: *Wie moeten er betrokken zijn?* Bij nieuwe systeemtechnologieën hebben grote bedrijven en overheden de middelen en belangen om vroege gebruikers van innovaties te zijn. Partijen in het maatschappelijk middenveld raken hierbij doorgaans pas later betrokken. Op die manier verscherpen deze nieuwe technologieën in eerste instantie bestaande maatschappelijke machtsverhoudingen. Ook draagt specifiek AI hieraan bij, doordat algoritmen op allerlei manieren minder welvarende burgers, etnische minderheden en vrouwen kunnen benadelen. In de huidige situatie is het vooral een groep technische specialisten die bij de ontwikkeling de vraagstukken rondom AI bespreekt (monoloog), terwijl dit ook een zaak zou moeten zijn van allerlei andere partijen en organisaties (dialoog). Bovendien is er vaak een grote afstand tussen de ontwikkelaars van AI-systemen en de maatschappelijke omgeving waarin die worden toegepast. Burgers en partijen in het maatschappelijk middenveld hebben expertise in te brengen en daarnaast hebben ze een belangrijke rol bij het geven van terugkoppeling over de werking van AI-systemen. Ook in deze zin draait het dus om dialoog.

Willen de organisaties die de rechten en belangen van burgers behartigen, hun rol bij AI kunnen waarmaken, dan moeten zij ook over de capaciteiten daartoe

kunnen beschikken. De ontwikkeling van eerdere systeemtechnologieën heeft laten zien dat de specifieke kennis van deze organisaties onmisbaar is bij de inbedding van dit soort technologieën, en daarmee bij de inbedding van AI in de samenleving. Toch ontbreekt hun stem momenteel in veel AI-discussies. In het bijzonder organisaties die zich richten op bijvoorbeeld de belangen van werknemers, patiënten, leraren, mensen die in armoede leven en achtergestelde en gediscrimineerde groepen, hebben nog onvoldoende kennis over AI. De WRR beveelt aan dat de overheid ze daarbij ondersteunt, via de subsidies die zij verstrekt en door trainingen en samenwerkingsverbanden te faciliteren. De formele en institutionele mechanismen voor belangenbehartiging, zoals ondernemingsraden en andere vormen van medezeggenschap, zijn hierbij bovendien beter te benutten.

Aanbeveling 5
Versterk de capaciteit van maatschappelijke organisaties om hun werk te verbreden naar het digitale domein, in het bijzonder met betrekking tot AI.

Ook is een goede terugkoppeling nodig tussen praktijk en tekentafel. Aandacht voor de kwaliteit en betrouwbaarheid van data, analysemethoden, en de werking en transparantie van AI-systemen is belangrijk. Evaluatie van de uitkomsten van AI-systemen en de terugkoppeling daarvan naar de ontwikkelaars en andere betrokkenen klinken als een logische vereiste voor AI-systemen, maar gebeuren in de praktijk onvoldoende. Een effectieve terugkoppeling is cruciaal voor het goed functioneren van AI-systemen en het beschermen van publieke waarden. Die terugkoppeling dient een dubbele afstand te overbruggen, namelijk allereerst tussen de ontwikkelaar en de gebruiker van het AI-systeem, en daarnaast met de personen die de effecten van het systeem ondervinden. Zowel de gebruiker als de aan het AI-systeem onderworpen personen zijn in een positie om fouten te herkennen, expertise in te brengen en verbeteringen voor te stellen. De WRR beveelt daarom aan dit proces van terugkoppeling te verbeteren; in de publieke sector, bij gemeenten, uitvoeringsinstanties en in domeinen waar beslissingen grote gevolgen voor burgers hebben, zou hiervoor een verplichte standaard moeten komen.

Aanbeveling 6
Draag zorg voor een goede terugkoppeling tussen de ontwikkelaar van AI, de gebruiker ervan en de personen die er in de praktijk de consequenties van ondervinden.

Regulering: Van reactie naar regie

De opgave 'regulering' speelt op het niveau van de samenleving als geheel en heeft als centrale vraag: *Wat voor kaders zijn nodig?* De afgelopen jaren zijn diverse reguleringsprocessen in gang gezet, met het voorstel voor een AI-Verordening van de EU als meest prominente uitdrukking. Deze reguleringsprocessen vinden overwegend plaats rondom acute en relatief helder afgebakende vraagstukken, zoals het gebruik van algoritmen voor fraudebestrijding, bias en discriminatie, transparantie en onbetrouwbare uitkomsten. Het debat over regulering richt zich hierbij overwegend op vragen over de relevantie van bestaande kaders dan wel de noodzaak van nieuwe regulerende en toezichthoudende instituten. Er is met andere woorden sprake van reactie. Naarmate AI meer ingebed raakt in de samenleving, ontstaan echter meer en meer raakvlakken met de publieke waarden waarover de overheid heeft te waken en treden ook tweede- en derde-orde-vraagstukken op die regulering behoeven. Immers, het zijn bij een systeemtechnologie behalve de concrete toepassingen vooral de bredere effecten ervan in de samenleving die vragen oproepen. De overheid dient zich naar het oordeel van de WRR voor te bereiden op de steeds grotere rol die hieruit volgt. Een verschuiving dus van reactie naar regie.

Het potentieel alomtegenwoordige karakter van AI maakt het lastig op voorhand te overzien welke kaders onder druk komen te staan of anderszins aanpassing behoeven. Toch mag de wetgever niet afwachten; daarvoor zijn de (publieke) belangen die op het spel staan te groot. Zaken als betrouwbaarheid, uitlegbaarheid en transparantie zijn absoluut belangrijk en vragen om regulerend handelen van de overheid, maar van doorslaggevend gewicht voor de verdere toekomst zijn uiteindelijk de inrichtingsvraagstukken waar een samenleving met AI voor komt te staan. Elektriciteit en de komst van de auto noodzaakten de wetgever om de bredere effecten daarvan op de samenleving te overwegen. Denk aan kabels boven of onder de grond en de aanleg van een wegenverkeersnet in afweging met de natuur. Bij de inbedding van AI spelen soortgelijke inrichtingsvraagstukken ten aanzien van de 'digitale leefomgeving', die ook zonder AI al zeer veel aspecten van de samenleving omvat. De WRR beveelt aan dat de wetgever op korte termijn een actievere rol pakt, ontwikkelingen veel meer vanuit een integraal perspectief adresseert en tijdig via regulering stuurt op de bredere economische en maatschappelijke context waarbinnen AI tot wasdom komt.

Aanbeveling 7
Koppel de regulering van AI aan een discussie over de inrichting van de digitale leefomgeving en stel een brede wetgevingsagenda op.

Cruciaal hierbij is dat de wetgever zich ook richt op de regulering van de maatschappelijke dynamiek waarmee AI gepaard gaat. Het gaat daarbij om de toename van surveillance en daarmee dataverzameling in de samenleving, de toenemende afhankelijkheid van de publieke sector van het bedrijfsleven in het digitale domein, en de machtsconcentratie bij grote bedrijven. De wetgever zal greep op de ontwikkelingen moeten blijven houden, anders verliest hij het vermogen om die ontwikkelingen op tijd te kunnen bijsturen. Hiertoe zal de overheid niet alleen moeten investeren in onderzoek en signalering, zoals nu gebeurt door bijvoorbeeld toezichthouders, maar ook concrete stappen moeten durven zetten. Wacht de overheid hiermee te lang, dan zal de dynamiek van de inbedding van AI de wetgever inhalen en hebben bepaalde spelers de macht inmiddels zo naar zich toegetrokken dat een weg terug nauwelijks nog mogelijk is. Op dat moment verliezen de bestaande kaders aan geldingskracht en komt de inrichting van de samenleving op basis van publieke waarden onder druk te staan.

Aanbeveling 8
Stuur via wetgeving actief op ontwikkelingen rondom surveillance en dataverzameling, de scheve verhouding tussen publiek en privaat in het digitale domein en machtsconcentratie.

Positionering: Van natie naar netwerk

De opgave van positionering heeft betrekking op het internationale toneel en betreft de vraag: *Hoe verhouden wij ons internationaal?* Het gaat hierbij over de rol die een nieuwe systeemtechnologie speelt bij het stimuleren van de competitiviteit van landen en de invloed ervan op de aard en uitkomsten van internationale conflicten. Als gevolg van deze twee dynamieken ontstaat doorgaans het idee van een mondiale race rondom de nieuwe technologie en pogen sommige landen zelfs om die volledig binnen de eigen landsgrenzen te ontwikkelen en te behouden. Dit speelt ook rondom AI. Allereerst wordt er veel gesproken over een 'AI-race', met de VS en China als koplopers. Talloze landen hebben de afgelopen periode AI-strategieën ontwikkeld om via een 'digitaal industriebeleid' aan die race mee te doen en het verdienvermogen met AI te versterken. Daarnaast groeit op het gebied van conflict en veiligheid het besef van de impact van AI, met autonome wapens als prominente toepassingen daarvan. Als Nederland zich onvoldoende positioneert op het gebied van AI, zal het kansen missen om met AI het verdienvermogen van ons land te versterken en zal het onvoldoende voorbereid zijn op de nieuwe veiligheidsrisico's die met AI gemoeid zijn.

De WRR bepleit dat we het idee dat Nederland verwikkeld is in een competitie met andere landen om welvaart en macht, minder centraal stellen en ons richten op de verbindingen die ons land met het buitenland heeft. Een focus op de natie moet dus plaatsmaken voor de focus op een netwerk. Internationale samenwerking kan plaatsvinden op verschillende gebieden, waaronder fundamenteel onderzoek, het opzetten van nieuwe diensten, zoals Gaia-X, coördinatie rondom bedrijven en wet- en regelgeving. Bijzondere aandacht verdient het proces van standaardisatie. Standaarden zijn immens invloedrijk en hebben grote effecten op de competitiviteit van landen. De WRR wijst er nadrukkelijk op dat standaardisatie in toenemende mate onderhevig is aan 'geopolitisering'. De EU en Nederland daarbinnen dienen hierop zeer alert te zijn en de samenwerking te zoeken met landen die dezelfde waarden onderschrijven. Om op al deze domeinen weloverwogen keuzes te kunnen maken, adviseert de WRR een integrale AI-diplomatie te ontwikkelen.

Aanbeveling 9
Versterk het Nederlandse verdienvermogen met een 'AI-diplomatie' die gericht is op internationale samenwerkingsverbanden, in het bijzonder binnen de EU.

De transitie van natie naar netwerk impliceert dat veiligheid niet alleen een zaak is van externe dreigingen aan de landsgrenzen, maar ook samenhangt met technologieën die burgers in hun dagelijkse leven gebruiken. Socialemediaplatformen, sensoren in de infrastructuur, besturingssystemen en communicatiesystemen en andere 'vernetwerkte' domeinen zijn allemaal potentiële kwetsbaarheden. Autoritaire landen als China en Rusland beschouwen digitalisering en AI als middelen voor de nationale veiligheid en hebben grote capaciteiten opgebouwd. Bovendien exporteren ze die technologieën ook. Nederland zal beter in kaart moeten brengen op welke middelen buitenlandse mogendheden inzetten en hoe dat ons democratische bestel onder druk kan zetten. Vervolgens dienen wij onze capaciteiten te versterken om dit tegen te gaan. Hoewel onduidelijk is met welke middelen de zogenoemde informatieoorlog te winnen valt, is er geen tijd te verliezen bij de opbouw van expertise en eventuele beleidskeuzes. Een snelle eerste stap die Nederland op dit gebied kan zetten, is aandacht hiervoor in het jaarlijkse cybersecuritybeeld.

Aanbeveling 10
Weet je als land ook in het AI-tijdperk te verdedigen; versterk daarom de Nederlandse capaciteiten tegen de groeiende 'informatieoorlog' en de export van digitale dictatuur.

Tot slot: een beleidsinfrastructuur voor AI

Ter ondersteuning van het werk dat nodig is om de inbedding van AI in de samenleving te ondersteunen en daarmee de vijf hiervoor geschetste opgaven ter hand te nemen, adviseert de WRR tot slot de opbouw van een beleidsinfrastructuur. AI zal een variëteit aan zowel sectorspecifieke als generieke publieke waarden raken. Met de tijd zullen de risico's maar ook de kansen voor die waarden scherper in zicht komen. Ook zal steeds vaker debat nodig zijn over de doelen die wij als samenleving willen nastreven en de vraag waar, waarvoor en onder welke condities we AI willen gebruiken. Bovendien vraagt AI om internationale samenwerking tussen landen, in het bijzonder binnen de EU. De WRR stelt vast dat het strategische belang van AI zowel in eigen land als mondiaal steeds meer wordt onderkend. Ook dat vraagt om een actieve overheidsrol.

Een beleidsinfrastructuur is in ieder geval nodig om de aankomende Europese AI-Verordening uit te voeren. Deze stelt dat de lidstaten één of meer nationale bevoegde autoriteiten aanwijzen om toezicht te houden op de toepassing en uitvoering van AI en als officieel contactpunt voor het publiek en andere actoren. Maar de opbouw van een beleidsinfrastructuur voor AI dient méér te omvatten, zo meent de WRR. Daarbij kan ons land zich laten inspireren door andere landen, die inmiddels AI-adviesraden en ambtelijke AI-bureaus instelden. De WRR acht het in deze fase te vroeg om een apart ministerie of een specifieke toezichthouder voor AI te bepleiten. Dit betekent echter niet dat de WRR de huidige status quo van het overheidsbeleid rondom AI als adequaat beoordeelt. Als eerstvolgende stap in de opbouw van een beleidsinfrastructuur bepleit de WRR daarom een coördinatiecentrum voor AI, dat aan beleidsdirecties, toezichthouders en uitvoeringsorganisaties een structuur biedt om regelmatig en rond uiteenlopende kwesties met elkaar in contact te treden en van elkaar te leren.

Slotaanbeveling
Bouw een beleidsinfrastructuur voor AI op, te beginnen met een AI-coördinatiecentrum voorzien van politieke verankering middels een ministeriële onderraad.

Het coördinatiecentrum kan kennis bij elkaar brengen, richting aanbrengen in de voor de overheid relevante vraagstukken, kansen en risico's rondom AI identificeren en een belangrijke coördinerende en faciliterende rol vervullen bij het opstellen van de bredere wetgevingsagenda die de WRR bepleit. De ervaringen hiermee kunnen in een volgende fase de basis vormen voor het faciliteren van de beleidsvoorbereiding en wellicht ook de beleidsbepaling en -uitvoering. De WRR acht het daarom wezenlijk dat het centrum een politieke verankering kent, zodat er snel beleid kan worden gemaakt als dat nodig is en

daartoe de politieke afstemming en sturing voorhanden zijn. Hiertoe beveelt de WRR de regering aan een ministeriële onderraad in te stellen waar zwaarwegende kwesties rondom digitalisering die om een integrale afstemming vragen, aan de orde komen.

Ten geleide

Dit rapport is opgesteld door een projectgroep bestaande uit prof.mr. Corien Prins (eerstverantwoordelijk raadslid), prof.dr. Haroon Sheikh (projectcoördinator), dr. Erik Schrijvers (projectcoördinator), drs. Eline de Jong (staflid), mr. Monique Steijns (staflid) en prof.dr.mr. Mark Bovens (raadslid). Dr.mr. Reijer Passchier (ex-staflid) was gedurende 2019 en een deel van 2020 verbonden aan de projectgroep. Eveneens tijdelijk verbonden aan de projectgroep waren Ivo Knottnerus (rijkstrainee), Julia Stroop, Tycho Tax, Mirthe Dankloff, Hanneke Roodbeen, Bart Gulden, Jord Goudsmit, Jurgen Timmerman, Nina Serpenti en Tessel van Oirsouw (stagiairs), en Riccardo Rapparini. Stafleden dr. Robert Went en prof.dr. Huub Dijstelbloem hebben bij aanvang van het project gediend als gesprekspartners. Ondersteuning werd verzorgd door Magda de Wit, Caroline Buser, Mitra Javanmardi en Dmitri Berkhout.

Opgave AI. De nieuwe systeemtechnologie kwam tot stand op basis van een uitvoerige studie van wetenschappelijke literatuur en beleidsstukken, interviews en discussiebijeenkomsten alsmede eigen analyse.

Verschillende onderzoekers hebben ter voorbereiding van dit rapport in opdracht van de WRR working papers opgesteld:

- Bennie Mols, *Internationaal AI-beleid. Domme data, slimme computers en wijze mensen.*
- Sjoerd Bakker en Pim Korsten (Freedomlab), *Artificiële intelligentie als een general purpose technology. Strategische belangen en verantwoorde inzet in historisch perspectief.*
- Babette Bakker, Devin Diran, Claudio Lazo, Govert Gijsbers en Amber Geurts (TNO), *Het technologisch ecosysteem van AI in Nederland.*
- Ernst Hirsch Ballin, *Mensenrechten als ijkpunten van artificiële intelligentie.*
- Monique Steijns, *AI van Repliek Gediend. Een verkenning van tegenmacht vanuit maatschappelijke organisaties.*

Deze working papers zijn beschikbaar op de WRR-website.

De interviews en gesprekken hebben we gevoerd met ruim 170 externe deskundigen in de publieke en private sector en in binnen- en buitenland. Gesproken is onder andere met gemeenten, toezichthouders, hoge colleges van staat, wetenschappers, vertegenwoordigers van bedrijven, organisaties in het maatschappelijk middenveld en de Nederlandse AI Coalitie. Er zijn diverse werkbezoeken afgelegd, onder andere aan de NAVO, France Stratégie en de Franse overheid, en

er is meerdere malen contact geweest met leden van de Eerste en Tweede Kamer en de interdepartementale werkgroep AI. Ook is een startbijeenkomst georganiseerd over AI en publieke waarden, en een bijeenkomst met de voorzitters van de adviesraden. De gesprekspartners zijn we zeer erkentelijk voor hun bijdrage aan dit rapport. Hun namen staan achterin dit rapport vermeld.

In de laatste fase van het project zijn teksten voorgelegd aan prof.dr. Luc Steels (emeritus hoogleraar Artificiële Intelligentie, Vrije Universiteit Brussel), Marleen Stikker en Tom Demeyer (respectievelijk directeur en CTO van Waag), prof.dr.mr. Stavros Zouridis (raadslid Onderzoeksraad voor Veiligheid), prof. dr. José van Dijck (universiteitshoogleraar Mediawetenschappen, Universiteit Utrecht) en prof.dr. Koen Frenken (hoogleraar Innovation Studies, Universiteit Utrecht). We danken hen voor hun commentaar en waardevolle suggesties.

Deel 1

Bouwstenen: introductie en duiding van AI als een nieuwe systeemtechnologie, vergelijkbaar met elektriciteit en de verbrandingsmotor

Inleiding

AI op een keerpunt

> "Een robot schreef dit hele artikel. Ben je al bang, mens? Ik ben geen mens. Ik ben een robot. Een denkende robot. Ik gebruik slechts 0,12 procent van mijn cognitieve capaciteit."

Dit rapport van de Wetenschappelijke Raad voor het Regeringsbeleid (WRR) is volledig door mensen geschreven. Wij verwachten ook dat het opstellen van dit type adviezen mensenwerk zal blijven. In contrast met wat het bovenstaande citaat aangeeft, geldt dat eveneens voor de journalistiek. Zoals later bleek, hadden mensen namelijk een flink aandeel in het artikel uit *The Guardian* van 8 september 2020 dat met de bovenstaande zinnen begon. Maar de aandacht die de publicatie genereerde, maakt wel iets duidelijk: artificiële intelligentie (AI) is tegenwoordig voorpaginanieuws.

De term *artificial intelligence* werd in de jaren vijftig van de twintigste eeuw bedacht en sindsdien werken wetenschappers aan de ontwikkeling van systemen die, met enige graad van autonomie, taken kunnen uitvoeren waarvoor intelligente vaardigheden vereist zijn. De laatste paar jaar is er echter iets veranderd. Waar AI vroeger het domein was van wetenschappers, geïnteresseerden en sciencefictionliefhebbers, spreekt de technologie nu breed tot de verbeelding. AI lijkt met andere woorden een vlucht te nemen, wat onherroepelijk effecten heeft op de samenleving. Een kleine greep van berichten van de laatste paar jaar:

2016

Google's programma AlphaGo verslaat de titelverdediger Lee Sedol in het spel Go. Toen schaakkampioen Garry Kasparov in de jaren negentig werd verslagen door IBM's Deep Blue, was de verwachting dat het nog een eeuw zou duren voordat een computer ook het complexere spel Go van de mens zou kunnen winnen.

Microsoft brengt de AI-bot Tay uit, die leert van menselijk gedrag op sociale media. Binnen een paar uur verandert Tay in een kwaadaardige trol met hatelijke opmerkingen over vrouwen en fascistische tweets.

2017

Berichten verspreiden zich dat AI-programma's van Facebook een eigen taal zouden hebben ontwikkeld die mensen niet begrijpen. Dit appelleert direct aan beelden van oncontroleerbare AI en de programma's zijn daarom snel afgesloten.

Ook spreekt de robot Sophia, gemaakt door Hanson Robotics, op een conferentie in Saoedi-Arabië en wordt haar staatsburgerschap verleend.

2018

CEO Sundar Pichai demonstreert Google Duplex, een AI-assistent die onder andere restaurantreserveringen kan doen en qua spraak niet van mensen te onderscheiden zou zijn.

Er verschijnt een deepfakevideo van president Barack Obama waarin het lijkt alsof hij een tekst uitspreekt die achter de schermen door komiek Jordan Peele wordt ingesproken.

2019

IBM's Project Debater neemt het op tegen een van 's werelds beste debaters, Harish Natarajan, in een debat over het subsidiëren van kleuterscholen. Na een argumentatieve krachtmeting tussen mens en machine, kent de jury Natarajan de winst toe.

2020

The Guardian publiceert een essay dat geschreven is door GPT-3, een taalgenerator van OpenAI. In het artikel betoogt GPT-3 dat de mens zich niet bedreigd hoeft te voelen door AI.

BostonDynamics publiceert een video waarin robots dansen op The Contours' *Do you love me*?

Het moge duidelijk zijn dat grote bedrijven flink investeren in AI en dat dat resultaten oplevert. De technologie raakt ingebed in het dagelijks leven van burgers via zoekopdrachten bij Google, de tijdlijn van Facebook, Siri – de digitale assistent van Apple – en de aanbevelingen van Amazon en Netflix. Ook heel veel Nederlandse bedrijven, van ASML tot Philips en ING tot Schiphol, gebruiken AI om diensten te personaliseren, producten te vernieuwen en hun bedrijfsprocessen te optimaliseren. Maar het momentum van AI betreft niet alleen maar bedrijven.

Ook overheden hebben hun vizier op AI gericht. Tientallen landen presenteerden de afgelopen jaren kort na elkaar hun nationale AI-strategieën. Sinds oktober 2019, toen staatssecretaris Mona Keijzer het Strategisch Actieplan voor AI (SAPAI) presenteerde, behoort ook Nederland tot die landen. Overheden zijn inmiddels grote gebruikers van AI en dat geldt ook voor de Nederlandse publieke sector. De politie, defensie en de douane zetten de technologie bijvoorbeeld in voor veiligheid, ziekenhuizen voor ondersteuning in de zorg, en Rijkswaterstaat om de publieke ruimte te verbeteren, net als veel gemeenten met hun 'smart city'-projecten.

Naast bedrijven en overheden heeft ook de populaire cultuur AI omarmd, vooral als een bron van dystopische toekomstscenario's. Al heel lang worden films gemaakt met kwaadaardige computersystemen zoals *Colossus: The Forbin Project* uit 1968 en *The Terminator* uit 1984. De laatste paar jaar herleeft de aandacht voor de toekomst die we tegemoet gaan met steeds slimmere computers, met films en series als *The Matrix, I Robot, Her, Ex Machina, Artificial Intelligence, Transcendence, Next, Black Mirror* en *West World*.

Los van deze fictie over een dystopische toekomst, is er ook breder in de samenleving aandacht voor actuele controverses rondom het gebruik van AI. Sociale bewegingen adresseren risico's en concrete misstanden. Op het gebied van defensie gaat de discussie bijvoorbeeld over drones die automatisch doelen kunnen identificeren en uitschakelen, de '*lethal autonomous weapons systems*', of kort – en weerzinwekkender – '*killer robots*'. In 2015 schreef een grote groep wetenschappers een open brief aan de Verenigde Naties om deze wapens te verbieden. In 2017 volgde een brief die ook door veel oprichters van bedrijven in het veld was ondertekend.

Een andere toepassing die veel losmaakt in de maatschappij, betreft zelfrijdende auto's. In 2016 was Joshua D. Brown de eerste persoon die in een zelfrijdende auto verongelukte en inmiddels zijn er al verschillende fatale ongelukken geweest met auto's van Uber en Tesla. Weer een andere AI-toepassing die veel controverse oproept, is gezichtsherkenning, waarbij gezichten door middel van *computer vision* in camerabeelden onderscheiden kunnen worden. Het

gevaar van totalitaire surveillance heeft tot een roep geleid om deze toepassing te verbieden. Verschillende Amerikaanse steden waaronder San Francisco, Boston en Portland, hebben gezichtsherkenning inmiddels gereguleerd of juist verboden. De Europese Commissie legt het gebruik van deze technologie strikt aan banden in de door haar voorgestelde AI-Verordening. Recente voorbeelden van controverse over AI in eigen land zijn het door de rechter verboden Systeem Risico Indicatie (SyRI) dat bedoeld was om fraude op te sporen, of het gebruik van algoritmes bij de toeslagenaffaire.

AI verlaat het lab en treedt de samenleving binnen

AI staat dus op een keerpunt. De technologie maakt definitief de overstap naar ons dagelijks leven en doet daarmee een hoop stof opwaaien. We kunnen dat keerpunt als volgt karakteriseren: *AI treedt vanuit het laboratorium nu de samenleving binnen* (zie figuur I.1). Dat is natuurlijk een simplificatie van de werkelijkheid. Laboratorium, de ruimte van onderzoek, en samenleving zijn niet radicaal van elkaar te scheiden. Laboratoria zijn ook een onderdeel van de samenleving en ideeën, mensen en praktijken bewegen continu heen en weer tussen beide.[1] Het laboratorium is bovendien geen vaststaande entiteit. Het laboratorium van Louis Pasteur is immers niet te vergelijken met de computerlabs uit de tijd van de Koude Oorlog of met de hedendaagse mondiale onderzoeksinstellingen. Toch is de beweging van lab naar samenleving een zinvolle duiding om het huidige momentum op het gebied van AI duidelijk te maken.

De oorsprong van het vakgebied is te herleiden tot een onderzoeksprogramma aan Dartmouth College in de Verenigde Staten in 1956. Al lang daarvoor werd over AI gefantaseerd, maar met dit programma startte het systematische onderzoek in het lab. Gedurende de daaropvolgende decennia hebben verschillende vormen van AI uit dat lab hun weg naar de samenleving gevonden. Programma's voor dammen en schaken bestaan al sinds de jaren zestig en beslisbomen worden al heel lang in allerlei digitale systemen toegepast. Vanaf de jaren tachtig groeit het gebruik van zogenoemde 'expertsystemen', programma's die worden gevoed met bijvoorbeeld medische kennis om artsen in hun werk te ondersteunen. Vanaf het begin waren er opzienbarende experimenten en demonstraties die tot de verbeelding van het bredere publiek spraken. Toch bleef de daadwerkelijke invloed van AI op economie en samenleving relatief beperkt. Tot recent.

Pas in het afgelopen decennium is de transitie van lab naar samenleving in alle ernst op stoom gekomen. We zien dat AI een rol van betekenis in de samenleving gaat spelen. Naast de onderzoekers in het lab raken andere partijen met hun eigen belangen betrokken bij de ontwikkeling ervan. Dat geldt allereerst

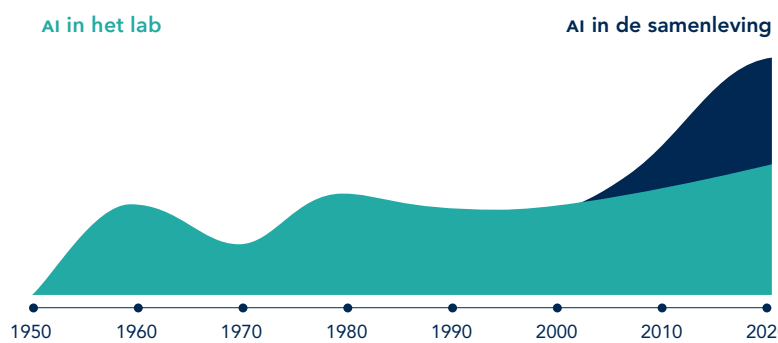
1 Zie bijvoorbeeld Latour 1983.

voor het bedrijfsleven. Een ijkpunt daarvoor was de overname van het Britse onderzoekslab DeepMind door Google in 2014. DeepMind is verantwoordelijk voor het eerdergenoemde programma AlphaGo dat in 2016 een menselijke kampioen versloeg in het spel Go. Grote technologiebedrijven zien in AI een belangrijke winstdrijver. Google en Microsoft beschrijven zichzelf als '*AI-first businesses*', bedrijven die van AI een prioriteit maken. Naast grote technologie-platformen groeit het aantal innovatieve start-ups, maar ook reeds bestaande bedrijven in andere sectoren richten zich in toenemende mate op AI.

Nationale AI-strategieën laten zien dat AI ook de interesse heeft gewekt van overheden. Zij zien de technologie als een belangrijke bron van toekomstige economische groei, een manier om hun eigen dienstverlening te verbeteren, maar ook als een fenomeen met risico's, waar regulering en toezicht voor ontwikkeld dienen te worden. Partijen in het maatschappelijk middenveld raken geëngageerd met AI wanneer zij opkomen voor benadeelde partijen, zich inzetten voor normatieve kaders of rechtszaken aanspannen om praktijken te toetsen. De onderzoekers uit het lab dragen naast hun technische expertise de laatste jaren ook bij aan normatieve discussies over de toepassingen van de technologie.

Ten slotte komt ook het bredere publiek met AI in aanraking. Niet alleen door de discussie die losbarst over allerlei toekomstbeelden, maar ook omdat zij de effecten van AI beginnen te merken. Bijvoorbeeld wanneer algoritmes een rol spelen in de diensten waar zij op rekenen – denk aan de toeslagenaffaire – of wanneer deze de aard van hun werk veranderen waardoor zij bijgeschoold moeten worden.

Figuur I.1 AI verlaat het lab en komt de samenleving in



Niemand weet hoe AI zich in de toekomst zal ontwikkelen. De invloed ervan op de samenleving wordt bovendien sterk bepaald door hoe de genoemde actoren ertegenaan kijken en op welke manier zij ermee omgaan. Al deze verschillende partijen hebben hun eigen belangen en waarden en hun eigen middelen om die belangen en waarden te behartigen. Hun belangen kunnen overeenkomen, bijvoorbeeld wanneer belangengroepen en media gedupeerde burgers steunen of wanneer overheden en bedrijven samenwerken om het nationale verdienvermogen te versterken. Partijen kunnen ook met elkaar botsen. Denk aan de spanning tussen de nadruk op openbaarheid van het wetenschappelijk onderzoek naar AI tegenover het belang van geheimhouding bij bedrijven. Of aan de spanning tussen burgers en overheden als het gaat over de inzet van surveillancetechnologie, waarbij veiligheid en privacy tegenover elkaar kunnen staan.

Een inzet van AI in de samenleving die in lijn is met onze publieke waarden, vraagt om samenwerking, onderhandeling, gewenning, debat en strijd. Anders geformuleerd, de intrede van AI vraagt om een complex proces *van inbedding in de samenleving*. Hoe kunnen wij dat proces het beste begeleiden en daar waar noodzakelijk proberen te beïnvloeden? Twee zaken moeten daarvoor nader onderzocht worden: de aard van AI als technologie en de relatie ervan tot publieke waarden.

Technologie en publieke waarden

De WRR schrijft dit rapport in reactie op een adviesaanvraag van de regering. In die adviesaanvraag staat: "Gelet op het toenemende gebruik van AI in nagenoeg alle sectoren (…) is er behoefte aan discipline-overstijgend onderzoek naar de impact van AI op publieke waarden." De regering verwijst daarbij naar zowel de eigenschappen van AI zelf, vragen rond de controle en aansprakelijkheid als bestaande problemen die met digitalisering te maken hebben, zoals privacy, cybersecurity en machtsconcentratie bij technologiebedrijven. Onderzoek vergt dus een analyse van AI en van publieke waarden.

In dit rapport bespreken wij wat AI is en hoe de technologie te karakteriseren valt. Over de impact van AI-toepassingen in allerlei domeinen bestaat inmiddels enorm veel literatuur. Om echter tot een overstijgend onderzoek van de impact van AI te komen, moeten wij wat afstand nemen en de vraag stellen over wat voor soort technologie we het hier hebben.

Bijzonder aan AI is hoe breed toepasbaar de technologie is. In de wetenschappelijke literatuur worden zeer breed toepasbare technologieën aangeduid als *general purpose technologies*. AI kunnen we zo vergelijken met eerdere technologieën als de stoommachine, elektriciteit en de verbrandingsmotor. In dit rapport redeneren we daarom vanuit analogieën met eerdere technologieën. De term die we kiezen om de aard van AI te begrijpen, is *systeemtechnologie*.

Met het woord 'systeem' doelen we zowel op de diverse set aan technologieën waaruit AI bestaat en waarmee zij verbonden is, als op het systemische effect dat ze op de samenleving heeft.

De karakterisering van AI als systeemtechnologie heeft meteen implicaties voor de benadering van het tweede element uit de adviesaanvraag: publieke waarden. Ook over de invloed van AI op publieke waarden is inmiddels een flinke hoeveelheid literatuur beschikbaar.[2] Daarnaast zijn op dit gebied talloze principes en kaders opgesteld. Een recente inventarisatie noemt meer dan driehonderd sets van ethische principes en richtlijnen rondom AI.[3] Belangrijk daarbij zijn die van de High-Level Expert Group on AI (AI HLEG) van de Europese Commissie (EC), van UNESCO en de toolkit van het AI Now Institute. Een aantal specifieke publieke waarden staat in veel van deze publicaties centraal. Meestal gaat het dan om uitlegbaarheid, transparantie, non-discriminatie, privacy, autonomie en aansprakelijkheid. In toenemende mate wordt ook in Nederland het publieke debat over deze waarden gevoerd. Dat is belangrijk en we zullen daar in dit rapport de nodige aandacht aan besteden. Tegelijkertijd is het risicovol om de impact van AI te willen vatten in een concrete lijst met publieke waarden.[4] Dat strookt immers niet met het dynamische karakter van de intrede van AI in de samenleving.

Als AI een systeemtechnologie is, zoals wij in dit rapport betogen, dan zal de impact ervan op publieke waarden nooit door middel van een lijst geïnventariseerd kunnen worden. Hiervoor zijn verschillende argumenten. Om te beginnen zal AI als systeemtechnologie namelijk in toenemende mate in de hele samenleving toegepast worden. Omdat we nu pas aan het begin van die ontwikkeling staan, heeft elke lijst een voorlopig karakter. Ook zal de technologie naast de bovengenoemde 'AI-specifieke' waarden ook invloed hebben op de waarden die centraal staan in de specifieke context waarin ze toegepast wordt. De adviesaanvraag van de regering gaat eveneens uit van die brede impact. Naast waarden als gelijke behandeling en keuzevrijheid, worden waarden als veiligheid, gezondheid en goede besluitvorming genoemd: waarden die in het spel komen als AI wordt toegepast in het verkeer, de gezondheidszorg of bij de overheid. Als AI in een bepaalde context toegepast kan worden, dan heeft ze potentieel invloed op alle publieke waarden die in die context van belang zijn.

---

2 Zie onder andere Vetzo et al. 2018; Kulk en Van Deursen 2020; Schermer et al. 2020.

3 Russell 2019: 249.

4 Zie ook het WRR Working Paper van Ernst Hirsch Ballin over mensenrechten als ijkpunten van artificiële intelligentie (Hirsch Ballin 2021). Mensenrechten worden daarin niet nagelopen om te beoordelen of op AI gebaseerde praktijken door de beugel kunnen. Hirsch Ballin maakt daarentegen duidelijk hoe AI aan de verwezenlijking van die rechten kan bijdragen.

Het effect van AI op publieke waarden is behalve breed ook onvoorspelbaar, zo leert de geschiedenis van systeemtechnologieën. Treinen en auto's veranderden niet alleen de mobiliteit, maar ook de inrichting van steden. Wonen en werken hoefden immers niet meer in dezelfde omgeving plaats te vinden. Of denk aan de invloed van elektrische apparatuur in het huishouden op de positie van de vrouw. Verwachtingen kunnen ook misplaatst zijn. Zo werd van de auto aanvankelijk verwacht dat die de stad schoner zou maken, omdat deze paarden met hun mest en bijbehorende ziekten uit het straatbeeld zou laten verdwijnen.[5]

Hiernaast speelt dat systeemtechnologieën de invulling van publieke waarden mede vormgeven. De auto maakte verre reizen en nieuwe vormen van jeugdcultuur mogelijk en beïnvloedde zo waarden als privacy, vrijheid en autonomie.[6] De impact van een technologie als AI op de publieke waarden is dus verre van eenduidig. De inventarisaties die nu gemaakt worden, zijn van groot belang, want zij brengen in kaart wat er nu allemaal gebeurt en geven zo houvast aan de discussie. Het risico is echter dat daarmee de suggestie kan worden gewekt van volledigheid en de illusie dat als voor die waarden zorg wordt gedragen, de impact van AI beheersbaar wordt.

Ten slotte is belangrijk dat het idee van 'impact' een misleidende voorstelling van zaken is. Als de samenleving en bijbehorende publieke waarden als iets statisch worden begrepen, zien we AI al snel als een extern fenomeen dat daaraan afbreuk kan doen. Die framing is veel te vinden in het debat over AI. We verliezen daarmee uit het oog dat AI de samenleving ook ten goede kan veranderen. Bijvoorbeeld door bepaalde publieke waarden sterker te bevorderen dan nu gebeurt. Hiervoor is een benadering nodig die het dynamische karakter van de inbedding van AI in de samenleving recht doet. Binnen een dergelijke benadering is de impact van AI niet het resultaat van externe druk maar de uitkomst van tweerichtingsverkeer met de samenleving.

Een historisch perspectief

Een onderzoek naar de inbedding van AI in de samenleving moet dus rekening houden met de breedte van het fenomeen, de onvoorspelbaarheid ervan, de wisselwerking tussen publieke waarden en technologie, en naast bedreigingen ook kansen voor publieke waarden identificeren. Hoe kan een dergelijk complex onderzoek aangevlogen worden op een manier die bruikbaar is voor het regeringsbeleid?

---

5 Gordon 2016.
6 Seo 2019.

Hiervoor gaan wij te rade bij het onderzoek naar de omgang van samenlevingen met grootschalige nieuwe technologieën en zoeken wij naar historische patronen. We gaan er niet van uit dat de geschiedenis zich herhaalt of dat technologie iets deterministisch is. In dit rapport besteden wij ook aandacht aan hoe AI anders is dan eerdere systeemtechnologieën. Tegelijkertijd menen we dat er in het verleden interessante patronen zijn te ontwaren die perspectief bieden op actuele vraagstukken. Een dergelijk langetermijnperspectief is daarom waardevol om inzicht te krijgen in het dynamische karakter van de maatschappelijke inbedding van systeemtechnologieën.

Op basis van onze studie van systeemtechnologieën onderscheiden wij in dit rapport *vijf maatschappelijke opgaven*. Het zijn breed geformuleerde opgaven die de blik richten op fundamentele aspecten van de samenleving, in het bijzonder van een samenleving die zich verknoopt met AI. Op die manier voorkomen we dat onze blik zich vernauwt tot concrete actuele dossiers of specifieke publieke waarden ten koste van structurele effecten en veranderingen. Met deze benadering van maatschappelijke opgaven beantwoorden we de vraag naar de impact die AI heeft op publieke waarden. Elke opgave toont namelijk een veelheid van publieke waarden die daarmee op het spel staat dan wel gemoeid is. Die publieke waarden bespreken we daarbij echter niet los van elkaar, maar vanuit hun onderlinge verbondenheid.

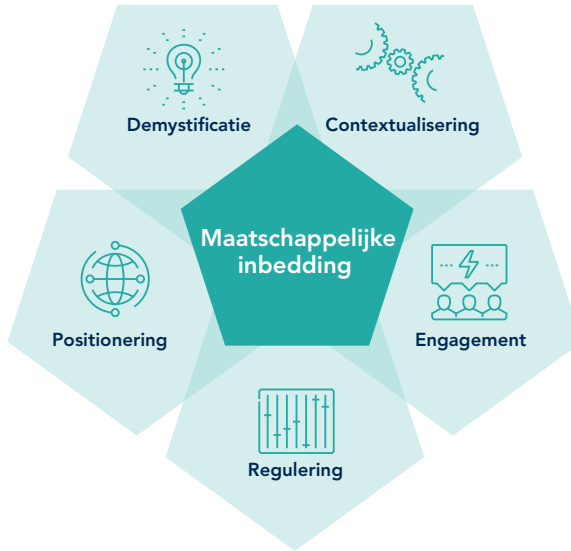
Maatschappelijke opgaven voor Nederland

We betogen in dit rapport dat de inbedding van een systeemtechnologie als AI bestaat uit de volgende vijf maatschappelijke opgaven:

1. **Demystificatie**
2. **Contextualisering**
3. **Engagement**
4. **Regulering**
5. **Positionering**

We zullen die opgaven individueel kort bespreken. Hun samenhang is echter cruciaal om het proces van maatschappelijke inbedding te begrijpen. De vijf opgaven hebben betrekking op verschillende niveaus en beantwoorden aan een vijftal centrale vragen. Demystificatie gaat over begrip op het niveau van AI als technologie en is gericht op de vraag: *Waar hebben we het over?* Contextualisering gaat over het niveau van de toepassing van een AI-systeem in een specifieke situatie en stelt de vraag: *Hoe gaat het werken?* Engagement heeft betrekking op de directe maatschappelijke omgeving van een AI-systeem en beantwoordt aan de vraag: *Wie moeten er betrokken zijn?* Regulering betreft het niveau van de samenleving als geheel, met als centrale vraag: *Wat voor kaders zijn nodig?* Positionering ten slotte gaat over het internationale niveau en laat ons nadenken over de vraag: *Hoe verhouden we ons internationaal?* Figuur I.2 visualiseert dit.

Figuur I.2 Vijf opgaven voor de maatschappelijke inbedding van AI



Deze opgaven hebben een universeel karakter. Ze waren van belang bij eerdere systeemtechnologieën als elektriciteit en de verbrandingsmotor en zijn wederom aan de orde bij de inbedding van AI. Ze betreffen namelijk maatschappelijke fundamenten als de publieke ruimte (demystificatie), bedrijfsvoering (contextualisering), interactie tussen maatschappelijke actoren (engagement), macht (regulering) en internationale betrekkingen (positionering).

Terwijl de opgaven zelf universeel zijn, is de invulling ervan afhankelijk van het type samenleving. Elke samenleving moet werken aan demystificatie, maar de aard en inrichting van de Nederlandse publieke ruimte en de actoren die daarbij een rol spelen, is anders dan in bijvoorbeeld de VS. Dat kan betekenen dat daar andere actoren invulling geven aan deze opgave dan in ons land. Of neem engagement. Elk land zal verschillende groepen van de bevolking bij een nieuwe technologie moeten betrekken. De rol van het maatschappelijk middenveld daarbij is anders in een samenleving als Nederland dan in een niet-democratisch land als China. De opgave van positionering gaat onder andere over veiligheid, dat essentieel is voor elke samenleving, maar per land zeer verschillend invulling krijgt.

Tezamen maken deze vijf opgaven het proces van inbedding van AI in de samenleving uit. Ze bieden handvatten om bij dat proces oog te hebben voor zaken die voor de Nederlandse samenleving essentieel zijn, zoals open debat, inspraak van betrokkenen, een overheid die kaders stelt en een veilige en welvarende

samenleving. Alhoewel wij de opgaven afzonderlijk bespreken, is het belangrijk om te benadrukken dat zij in de praktijk vaak nauw met elkaar verbonden zijn. Ze zijn niet los van elkaar of volgtijdelijk te begrijpen, maar moeten in hun onderlinge samenhang worden bezien.

Met deze benadering, gericht op maatschappelijke opgaven, beoogt de WRR bij te dragen aan een volgende stap in het publieke debat over AI. Dat debat barstte een klein decennium geleden los met grote toekomstverwachtingen. Visionaire auteurs voorzagen een wereld van zelfrijdende auto's, de genezing van allerlei ziekten en algoritmes die het werk van ons overnamen. Anderen waarschuwden juist voor een dystopische toekomst van machines die over mensen zullen heersen.

In de laatste paar jaar is het debat van aard en toon veranderd. Inmiddels worden AI-toepassingen op allerlei plaatsen geïmplementeerd en daarmee verschoof de focus van toekomstscenario's naar meer acute en actuele vraagstukken. Zo werd bijvoorbeeld duidelijk dat HRM-algoritmes vrouwen benadeelden en dat algoritmes gebruikt door veiligheidsdiensten mensen van kleur discrimineerden. Overheidsorganisaties overal ter wereld bleken op algoritmes te vertrouwen die zij amper begrepen of konden uitleggen. De toon van het debat in deze fase is dan ook in hoge mate negatief. Dat is niet vreemd. Zoals gezegd treedt AI de samenleving binnen, maar staan we pas aan het begin van het proces van inbedding. Het is net als de eerste auto's die de weg op gingen zonder veiligheidsgordels, airbags, verzekeringen, nummerplaten, verkeersregels of rijexamens. Of als het begin van de massaproductie van voedsel en medicijnen zonder veiligheidsstandaarden, bijsluiters, keurmerken of toezichthouders. Kortom: in deze fase zal er veel fout gaan en ontstaan allerlei misstanden door gebrek aan ervaring of duidelijke regels. Een groot risico is echter dat we door alle negatieve berichtgeving het zicht verliezen op de positieve bijdrage die AI aan de samenleving kan leveren. Of dat we dusdanig in beslag genomen worden door de risico's op de korte termijn dat we de vaak grotere gevolgen op de lange termijn uit het oog verliezen.

Mede daarom is het van belang dat we een volgende stap zetten in de discussie over AI en de impact van deze technologie structureel benaderen. Dat impliceert dat we ons niet alleen richten op acute vraagstukken en knelpunten, maar een gebalanceerd beeld ontwikkelen van de langetermijninbedding van AI in de samenleving. Daarbij spelen de vijf hiervoor genoemde maatschappelijke opgaven een wezenlijke rol. Wat houden deze opgaven in?

De vijf opgaven

De eerste opgave betreft de *demystificatie* van AI. Dat is een opgave die vooral ook het brede publiek betreft. Er doen veel mythes over AI de ronde. Die leiden

niet alleen tot een vertekend beeld van de technologie, maar ook tot overspannen verwachtingen enerzijds en angsten en afkeer anderzijds. Ondanks de aankondigingen van verschillende ondernemers en visionairs laat de zelfrijdende auto bijvoorbeeld al jaren op zich wachten. Een begrip van de immense uitdagingen voor AI op dit gebied maakt duidelijk waarom die verwachtingen onrealistisch waren. Omgekeerd zijn veel angsten over kwaadaardige AI die de wereld over zal nemen evenmin realistisch. Demystificatie gaat dus over een geïnformeerd beeld van wat de technologie op dit moment en de komende jaren wel en niet kan. Kortom: waar hebben we het over als het gaat om AI? We zullen zien dat er rondom AI mythen bestaan over de werking ervan, de toekomstige effecten en over digitale technologieën in het algemeen.

De tweede maatschappelijke opgave is de *contextualisering* van AI. Dit is een opgave voor alle actoren die AI in een specifiek domein toepassen en haar daar willen laten functioneren – de vraag: hoe gaat de technologie werken? Deze actoren kunnen bedrijven zijn, maar ook publieke instellingen. Contextualisering betreft in de eerste plaats het technische ecosysteem. Systeemtechnologieën kunnen alleen adequaat functioneren als er ook voldoende zorg is gedragen voor ondersteunende technologieën. Zoals de verbrandingsmotor de staalindustrie vooronderstelde, vragen AI-algoritmes om data, hardware en andere technologische ondersteuning. Het technische ecosysteem betreft daarnaast ook emergente technologieën: andere nieuwe technologieën die gelijktijdig opkomen, op elkaar kunnen inhaken en zo elkaar versterken. Denk bij AI aan het Internet of Things, blockchain en *quantum computing*. Contextualisering gaat daarnaast ook over niet-technische aspecten, zoals de opname van de nieuwe technologie in bedrijfsprocessen. Ook als een nieuwe technologie in het lab helemaal goed functioneert, betekent dat nog niet dat ze in de toepassing ook werkt. Het kost tijd om processen aan te passen, businessmodellen te ontwikkelen en mensen wegwijs te maken. Praktijk en technologie moeten zich aan elkaar aanpassen.

Maatschappelijke inbedding betekent in de derde plaats het *engagement* van belanghebbenden. De centrale vraag is hier: wie moeten er betrokken zijn? Wanneer AI namelijk in meer en meer contexten wordt toegepast, zullen steeds meer partijen in de samenleving erdoor geraakt worden en een belang hebben bij het in gebruik nemen van de technologie. Vooral actoren in het maatschappelijk middenveld spelen een centrale rol in het debat over een juiste toepassing van AI, maar ook individuele onderzoekers of bedrijven kunnen zich hiervoor inzetten.

Het belang van het engagement van dit soort partijen is groot, vooral in de vroege fasen van de ontwikkeling van een technologie omdat de effecten dan moeilijk te overzien zijn. Partijen in het maatschappelijk middenveld kunnen

al vroeg een agenderende en signalerende rol vervullen. Dat kunnen zij doen door misstanden aan te kaarten en aandacht te vragen voor de slachtoffers van ongevallen. Denk aan de fatale incidenten met zelfrijdende auto's of aan etnisch profileren door algoritmes. Geëngageerde stakeholders kunnen maatschappelijk zwakkere groepen of uitgesloten partijen een stem geven. Ook de (data) journalistiek kan die rol vervullen. In het verlengde hiervan ligt ook het maatschappelijk protest, dat in het verleden vaak tot betere en veiligere technologie heeft geleid. Behalve partijen in het maatschappelijk middenveld kunnen wetenschappers en technische experts, maar ook werknemers van technologiebedrijven en professionals wiens vakgebied door AI wordt beïnvloed, betrokken raken bij de technologie.

Maatschappelijke inbedding vraagt ten vierde om de *regulering* van AI. Vooral bij deze opgave heeft de overheid een belangrijke rol te spelen. Het dilemma daarbij is grofweg dat in een vroege fase een technologie redelijk goed is te reguleren, onder meer via reeds bestaande regels, maar dat er nog veel onduidelijkheid heerst over de positieve en negatieve effecten ervan. In een latere fase wordt duidelijker waar regulering nodig is, maar tegen die tijd zijn eerder genomen beslissingen en ontstane machtsstructuren vaak moeilijk te corrigeren. Dit dilemma is belangrijk, want we zien dat systeemtechnologieën veelal gepaard gaan met de opkomst van bedrijven die monopoliemacht hebben of andere overmatige invloed op burgers uitoefenen. Die macht dient geadresseerd te worden om democratisch gelegitimeerde keuzes over publieke waarden te waarborgen. Om de vraag 'wat voor kaders zijn nodig?' te kunnen beantwoorden is in eerste instantie goed zicht nodig op het noodzakelijke instrumentarium en de mate waarin bestaande regulering adequaat of in gebreke is. Bovendien is het bij de opgave van regulering belangrijk dat er naast aandacht voor acute vraagstukken ook voldoende aandacht is voor ontwikkelingen die de inbedding van AI onder druk zetten, zoals massasurveillance en de groeiende afhankelijkheid van private aanbieders in het digitale domein.

De vijfde en laatste opgave die we onderscheiden, is *positionering*. Hierbij gaat het om de vraag: hoe verhouden wij ons internationaal? Die vraag valt uiteen in twee verbonden zaken. Het gaat ten eerste over het verdienvermogen van Nederland. Om een welvarende en innovatieve samenleving te blijven, moet gekeken worden naar de AI-capaciteiten die Nederland heeft. Met het Strategisch Actieplan voor AI (SAPAI) heeft Nederland een AI-strategie. Vragen die hierbij horen zijn: speelt er een mondiale race om AI? Waar moet Nederland zich op richten en dient er een soort 'AI-diplomatie' ontwikkeld te worden? Naast verdienvermogen gaat de vraag hoe Nederland zich positioneert, ook over veiligheid. Veel aandacht gaat hierbij uit naar de gevaren van zogenoemde autonome wapens. De impact van AI op veiligheid is echter veel breder. AI beïnvloedt ook andere militaire domeinen en heeft bij uitstek invloed op

veiligheid in het civiele domein. Denk hierbij aan de groeiende informatieoorlog die online gaande is, maar ook aan de export van allerlei civiele technologieën, zoals slimme camera's, die voor dictatoriale doeleinden ingezet kunnen worden. Alhoewel verdienvermogen en veiligheid verschillende vraagstukken lijken, is het belangrijk op te merken dat ze internationaal steeds meer met elkaar verweven raken (geo-economie). Dat heeft implicaties voor de positionering van Nederland.

Opzet van het rapport

In dit rapport doet de WRR aan de hand van deze vijf opgaven van maatschappelijke inbedding aanbevelingen voor het Nederlandse regeringsbeleid. AI en de maatschappelijke inbedding daarvan zijn complexe en veelomvattende onderwerpen, die flink wat uitleg vergen. Dit rapport is daarom omvangrijk. Om de lezer daarin wegwijs te maken, hebben we de tekst in drie delen opgesplitst: deel 1 met bouwstenen voor het onderzoek; deel 2 met maatschappelijke opgaven; en deel 3 met conclusies en aanbevelingen. Lezers die meer willen weten over AI, adviseren we zich te richten op deel 1. Lezers die voornamelijk geïnteresseerd zijn in de maatschappelijke opgaven die de inbedding van AI met zich meebrengt, kunnen zich richten op deel 2. Ter ondersteuning hiervan is het wel belangrijk om in deel 1 de passages over de definitie van AI en de duiding ervan als systeemtechnologie door te nemen. Voor wie meteen weten wil hoe de overheid de maatschappelijke inbedding van AI mede vorm kan geven, om publieke waarden ook naar de toekomst toe te kunnen beschermen, is deel 3 de aangewezen stof om te lezen. Voor het overzicht staan in de verschillende hoofdstukken aparte kaders die de kernpunten samenvatten.

In het eerste deel, dat uit drie hoofdstukken bestaat, leggen wij het fundament voor het onderzoek naar de maatschappelijke inbedding van AI. Hoofdstuk 1 introduceert het thema van artificiële intelligentie: wat is AI, hoe valt de technologie te definiëren en welke keuze maken wij daarin? Vervolgens schetsen we de historische ontwikkeling van AI. Van vroege voorstellingen volgen wij het pad naar het begin van AI in het lab in 1956 en de verschillende 'golven' die de technologie sindsdien gekend heeft. Hoofdstuk 2 gaat over de recente ontwikkelingen rondom AI en laat zien hoe de technologie de laatste jaren vanuit het lab de samenleving ingaat. We behandelen daar de belangrijkste toepassingsgebieden, recente ontwikkelingen in het lab en de manier waarop het maatschappelijk debat over AI is losgebarsten. In hoofdstuk 3 duiden wij AI als technologie. Daarin kijken wij naar de literatuur over verschillende soorten technologieën en de manier waarop AI daarmee te vergelijken valt. We introduceren hier onze notie van AI als systeemtechnologie. Aan de hand van de historische inbedding van systeemtechnologieën lichten wij de vijf maatschappelijke opgaven in algemene zin toe.

In deel 2 werken we die vijf opgaven uit in relatie tot de inbedding van AI in de samenleving. Hoofdstukken 4 tot en met 8 bespreken elk een van de afzonderlijke opgaven en gaan over respectievelijk demystificatie, contextualisering, engagement, regulering en positionering. Deze hoofdstukken vormen de kern van onze analyse. In de hoofdstukken van deel 2 bespreken we wat de opgave voor AI betekent, wat de status daarvan momenteel in Nederland is en welke actoren erbij betrokken zijn.

In deel 3 ten slotte gaan we in op de implicaties die onze analyse heeft voor overheidsbeleid. Hoofdstuk 9 presenteert onze hoofdboodschap en resumeert de vijf opgaven van inbedding van AI waaraan we vervolgens onze aanbevelingen verbinden. Per opgave presenteren we twee aanbevelingen met bijbehorende concrete actiepunten om zo antwoord te geven op de adviesaanvraag van het kabinet over de impact van AI op publieke waarden. Ten slotte doen we nog een laatste aanbeveling, om ook antwoord te bieden op de tevens in de aanvraag gestelde vraag naar het instrumentarium van de overheid.

1. Artificiële intelligentie: definitie en achtergrond

1.1 Definities van AI

Om te kunnen onderzoeken wat ervoor nodig is om AI in te bedden in de samenleving, is een goed begrip van AI noodzakelijk. Kortom, wat verstaan we daaronder, hoe heeft de technologie zich ontwikkeld en waar staat ze nu?

Het blijkt niet eenvoudig om AI te definiëren. Het ontbreekt dan ook aan een algemeen geaccepteerde definitie van AI.[7] Tal van definities zijn in omloop, wat gemakkelijk tot verwarring leidt. Het is dan ook belangrijk om helderheid te creëren over het gebruik van de term. We bespreken hier verschillende definities van AI en geven vervolgens aan hoe we de term gebruiken. Dat er diverse definities in omloop zijn, zal geen kwestie van onzorgvuldigheid blijken, maar inherent aan het fenomeen van AI zelf.

In de meest brede definitie wordt AI gelijkgesteld met algoritmes. Voor onze analyse is dat echter een weinig bruikbare definitie. Algoritmes worden al langer benut en breder gebruikt dan uitsluitend binnen het veld van AI. De term 'algoritme' is afgeleid van de naam van de negende-eeuwse Perzische wiskundige Mohammed ibn Musa al-Kharizmi en verwijst naar een gespecificeerde instructie om een probleem op te lossen of een berekening uit te voeren. Definiëren we AI als het gebruik van algoritmes, dan vallen veel simpele zaken daar in principe onder, zoals de berekeningen van een rekenmachine en zelfs de instructies van een kookboek.

In de meest stringente definitie staat AI voor de nabootsing door computers van de intelligentie die eigen is aan mensen. Puristen wijzen erop dat veel van de huidige toepassingen nog relatief simpel en daarom geen *echte* AI zijn. Ook deze definitie is voor dit rapport daarom onvoldoende bruikbaar. Zouden we deze immers hanteren, dan impliceert dit dat er momenteel nog nergens sprake is van AI. Met andere woorden, we zouden het fenomeen in feite weg-definiëren.

Een gangbare definitie van AI is dat de technologie de nabootsing van diverse complexe menselijke vaardigheden door machines betreft. Veel levert deze definitie echter niet op. Eigenlijk zegt ze precies hetzelfde als de term 'kunstmatige intelligentie', alleen met andere woorden. Zolang die complexe menselijke vaardigheden niet gespecificeerd worden, blijft onduidelijk wat AI is. Hetzelfde geldt voor de definitie van AI als het uitvoeren door computers van complexe taken in complexe omgevingen.

---

7 Russell en Norvig 2020.

Andere definities proberen wel invulling te geven aan die vaardigheden en taken. De computerwetenschapper Nils John Nilsson spreekt bijvoorbeeld van een technologie die "adequaat kan functioneren door op een vooruitziende manier met de omgeving om te gaan".[8] Anderen spreken van de vermogens om waar te nemen, doelen na te streven, acties in gang te zetten en te leren van een feedbackloop.[9] Een vergelijkbare definitie is die van de AI HLEG van de Europese Commissie (EC): AI betreft "systemen die intelligent gedrag vertonen door hun omgeving te analyseren en – met enige graad van autonomie – actie te ondernemen om specifieke doelen te bereiken."[10]

Deze taakgerelateerde definities geven ons al een beter begrip van wat we onder AI kunnen verstaan. Toch kent ook deze uitleg zijn beperkingen. Termen als 'enige graad van autonomie' houden een zekere vaagheid in de definitie. Daarnaast lijken zij nog erg breed en zaken te omvatten waarvan de WRR niet geneigd is die AI te noemen. De definitie van Nilsson gaat bijvoorbeeld ook op voor een klassieke thermostaat. Ook die is in staat de omgeving waar te nemen (het meten van de temperatuur van de kamer), doelen na te streven (de geprogrammeerde temperatuur), acties in gang te zetten (de verwarmingsstand aanpassen) en te leren van een feedbackloop (stoppen wanneer de geprogrammeerde temperatuur bereikt is). Aan alle in de definitie geformuleerde voorwaarden wordt dus voldaan, maar de meeste mensen zullen niet geneigd zijn om een thermostaat als AI te beschouwen.

Dat het zo moeilijk is om AI helder af te bakenen, is overigens niet verrassend. Het betreft namelijk een imitatie of simulatie van iets dat we zelf nog niet volledig begrijpen: menselijke intelligentie. Daar wordt al lang onderzoek naar gedaan, onder andere door psychologen, gedragswetenschappers en neurologen. We weten veel van intelligentie en het menselijk brein, maar die kennis is verre van volledig en er is geen overeenstemming over wat menselijke intelligentie nu precies is. Zolang dat niet het geval is, is het onmogelijk om precies te zijn over hoe die intelligentie kunstmatig nagebootst kan worden.

Er is bovendien een sterke interactie tussen het onderzoek naar menselijke intelligentie enerzijds en dat naar kunstmatige intelligentie anderzijds. Immers, er is sprake van een co-evolutie waarbij dat wat als menselijke intelligentie wordt beschouwd, mee-ontwikkelt met AI. Dat illustreren we aan de hand van onderzoek naar schaken, iets dat AI sinds de jaren negentig op hoog niveau kan.

---

8 Nilsson 2009: 13.
9 Zie bijvoorbeeld DenkWerk 2018.
10 High-Level Expert Group on Artificial Intelligence 2019a. Aan het eind van het document updatet de AI HLEG de initiële definitie met een uitgebreide toelichting op de verschillende elementen ervan.

In de jaren vijftig voorspelde een expert: "Als een succesvolle schaakmachine gebouwd kan worden, dan bereiken we de kern van het menselijke intellectuele vermogen".[11] De Russische wiskundige Alexander Kronrod noemde schaken in 1965 "het fruitvliegje van intelligentie", oftewel de sleutel om haar te begrijpen.[12] Zo waren ook de reacties toen een computer er uiteindelijk in slaagde een schaakgrootmeester te verslaan. Toen in 1997 Garry Kasparov werd verslagen door Deep Blue, de schaakcomputer van IBM, stond op de voorpagina van *Newsweek*: "The brain's last stand". Schaken werd beschouwd als het summum van menselijke intelligentie. Dat is op het eerste gezicht ook niet zo gek omdat het spel voor mensen moeilijk is om te leren en zij die goed zijn in schaken vaak als zeer slim worden beschouwd. Vanuit dat idee meenden commentatoren dus dat de overwinning van Deep Blue een enorme doorbraak was op weg naar menselijke intelligentie en dat het binnen handbereik lag dat de computer de mens zou gaan overtreffen in allerlei activiteiten die we als makkelijker dan schaken beschouwen.

Dat is echter niet gebeurd. Inmiddels kijken we anders tegen deze vorm van intelligentie aan. Schaken is niet de kroon op de menselijke intelligentie. Het is een rekenkundig vraagstuk met heel duidelijke regels en een eindige set van mogelijkheden. Een schaakprogramma is eigenlijk niet heel anders dan een rekenmachine die ook dingen kan die zelfs voor heel slimme mensen te moeilijk zijn. Daarmee is het echter nog geen kunstmatige vorm van menselijke intelligentie.

Schaken leek lange tijd heel geavanceerd. Maar zoals jarenlang onderzoek heeft laten zien, is het herkennen van een kat op een foto – iets dat AI pas de laatste jaren heeft geleerd – veel complexer. Dit fenomeen is ook wel bekend komen te staan als de paradox van Moravec: bepaalde zaken die voor mensen heel moeilijk zijn, zoals schaken of geavanceerde calculus, zijn voor computers vrij gemakkelijk.[13] Maar zaken die voor ons mensen heel eenvoudig zijn, zoals de perceptie van objecten en motorische vaardigheden als het doen van de afwas, blijken voor computers juist heel moeilijk: "(…) it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility."[14]

---

11 Bostrom 2016: 14.
12 Floridi 2014: 139.
13 Moravec 1988. Een andere formulering komt van de AI-wetenschapper Donald Knuth. Hij merkte op dat AI erin slaagt om dingen te doen waar mensen over moeten nadenken, maar faalt bij taken die mensen zonder nadenken doen, zoals objecten herkennen, beelden analyseren en een arm bewegen (Bostrom 2016: 17).
14 Moravec 1988: 15.

Dit wijst op een patroon dat in de geschiedenis van AI steeds terugkeert: wat mensen als een complexe vorm van menselijke intelligentie beschouwen, evolueert mee met de vaardigheden van computers. Wat op een bepaald moment wordt beschouwd als een knap staaltje kunstmatige intelligentie, wordt met de tijd gezien als een simpele calculatie die de naam AI niet meer verdient. Pamela McCorduck noemt dit het 'AI-effect': zodra een computer erachter komt hoe iets te doen, is de reactie dat het 'slechts een berekening' betreft en geen daadwerkelijke intelligentie. Volgens Nick Bostrom, de directeur van het Oxford Institute for Internet Governance, is AI dan ook datgene wat wij op een bepaald moment indrukwekkend vinden. Zodra het dat niet meer is, noemen we het simpelweg software.[15] Dat geldt bijvoorbeeld voor een schaakapplicatie op de telefoon. De problemen om AI te definiëren zijn dus niet het gevolg van een tekortkoming of slordigheid, maar komen voort uit het feit dat wij niet voor langere tijd vast kunnen stellen wat de intelligentie die we kunstmatig willen nabootsen, precies inhoudt.

In dit verband wordt ook beweerd dat het gebruik van de aanduiding 'intelligentie' misleidend is. Dat gebruik suggereert ten onrechte dat machines hetzelfde doen als mensen. Sommigen stellen daarom voor om andere termen te gebruiken. Volgens Agrawal, Gans en Goldfarb brengt de huidige technologie ons geen intelligentie, maar slechts een component daarvan: voorspellingen. Zij spreken dan ook van '*prediction machines*'.[16] De filosoof Daniel Dennett gaat nog verder en stelt dat wij AI niet op mensen moeten modelleren. Het gaat niet om kunstmatige mensen, maar om een nieuw soort entiteit. Hij vergelijkt die entiteit met orakels: entiteiten die een voorspelling doen, maar in contrast met mensen geen persoonlijkheid, geweten of emoties hebben.[17] Met andere woorden, AI lijkt te doen wat mensen doen maar is in feite wat anders. Edsger Dijkstra illustreerde dit met de vraag: zwemmen onderzeeërs?[18] Wat deze vaartuigen doen, lijkt op wat bij mensen zwemmen heet, maar om dat zo te noemen is een vergissing. Evengoed doet AI dingen die lijken op de intelligente dingen die wij doen, maar in feite gaat ze heel anders te werk.

Dit perspectief werpt ook licht op de genoemde Moravec-paradox. Het herkennen van gezichten is voor mensen gemakkelijk, maar voor computers moeilijk. Dat komt omdat het herkennen van anderen in de evolutie zeer belangrijk was om te overleven en ons brein heeft zich aangeleerd dit te doen zonder daarover na te denken. Goed kunnen schaken was evolutionair niet essentieel en is

15 Bostrom 2016.
16 Agrawal et al. 2018: 2, 39. Gebaseerd op het werk van Jeff Hawkins zien deze auteurs 'voorspelling' als de basis van intelligentie.
17 Dennett 2019.
18 Dignum 2019.

daarom lastiger. Althans het vraagt om de nodige denkkracht. Omdat computers niet volgens deze biologische evolutie ontwikkeld zijn, liggen hun krachten ergens anders dan die van de mens. Belangrijk aan deze benadering is dat wij AI niet te zeer moeten begrijpen vanuit menselijke intelligentie. Desalniettemin: de term 'artificiële intelligentie' is inmiddels zo gangbaar geworden, dat het niet zinvol is om te proberen die te vervangen.

Tot slot wordt AI ook vaak gelijkgesteld aan de nieuwste technieken. Zoals we later zullen zien, is AI de laatste jaren in een stroomversnelling gekomen. Een belangrijke oorzaak daarvan is de vooruitgang in een specifiek onderdeel van het vakgebied, namelijk 'machine learning' (ML). De innovatie op dit terrein heeft geresulteerd in wat inmiddels 'deep learning' (DL) heet. Het is deze technologie die achter recente mijlpalen zit als het herkennen van gezichten door computers en het spelen van spellen als Go. In contrast met meer traditionele benaderingen waarbij computersystemen vaststaande regels toepassen, herkennen ML- en DL-algoritmes zelf patronen in de data. Daarom wordt in deze context ook gesproken van *zelflerende algoritmes*. Veel mensen die het tegenwoordig over AI hebben, doelen op dit type algoritmen en vaak specifiek op deep learning. De aandacht voor deze techniek is met name belangrijk omdat verschillende prangende vragen rondom AI, zoals problemen van uitlegbaarheid, in het bijzonder hier spelen.

Gegeven alle besproken en andere definities, kiezen wij ervoor AI open te definiëren. Twee overwegingen zijn hierbij relevant. Ten eerste: het is voor het doel van dit rapport onverstandig om de definitie van AI te beperken tot een specifiek onderdeel van de technologie. Zouden we ons bijvoorbeeld beperken tot het hiervoor besproken deep learning, dan negeren we dat veel van de vraagstukken, waaronder het zogenoemde 'black box'-vraagstuk, rondom AI evengoed spelen bij methoden in andere domeinen van AI, zoals logische systemen. Bovendien maken de meeste toepassingen van AI binnen de overheid geen gebruik van geavanceerde technieken als deep learning, terwijl daar wel degelijk allerlei belangrijke vraagstukken spelen die we in dit rapport aan de orde moeten stellen. Een te enge definitie zou die toepassingen buiten ons blikveld plaatsen.

De sprong voorwaarts is weliswaar groot gegeven de ontwikkelingen op het terrein van deep learning, maar – zoals wij aan het eind van dit hoofdstuk zullen toelichten – er wordt ook gewezen op verschillende tekortkomingen van deze techniek. De toekomstige vooruitgang in AI zou daarom weleens uit andere domeinen kunnen komen. Om daarvoor open te staan, is het verstandig om de nodige ruimte te laten in de definitie van AI.

Ten tweede: zoals hiervoor besproken, brengt de aard van het vakgebied met zich mee dat wat wij onder AI verstaan door de tijd heen verschuift. In plaats van AI te beschouwen als een discipline die helder af te bakenen is met duidelijke definities en vaste methodieken, kunnen we haar beter beschouwen als een complex en divers werkveld dat gericht is op een horizon. De stip op die horizon is het begrijpen en simuleren van alle menselijke intelligente vaardigheden. Dat doel wordt ook wel *Artificial General Intelligence* (AGI) genoemd. Andere benamingen ervoor zijn 'sterke AI'(*strong AI*) of 'volledige AI' (*full AI*). Het is echter nog maar de vraag of het punt op de horizon van een dergelijke generieke AI ooit bereikt zal worden. De meeste experts plaatsen het, indien het al uitkomt, ten minste een aantal decennia in de toekomst.[19]

Een vaststaand begrip van AI als de imitatie van volledig menselijke intelligentie is voor dit rapport weinig bruikbaar. We hebben een begrip nodig waarmee het hele palet aan toepassingen dat nu en in de nabije toekomst de weg naar de praktijk vindt, in beeld komt. De definitie van de AI HLEG biedt die ruimte. Deze omschrijving van AI als 'systemen die intelligent gedrag vertonen door hun omgeving te analyseren en – met enige graad van autonomie – actie te ondernemen om specifieke doelen te bereiken' omvat de toepassingen die wij op dit moment kwalificeren als AI en laat tegelijkertijd ruimte voor een toekomstige verschuiving van die kwalificatie. Onder deze definitie vallen naast de geavanceerde technieken van machine learning en deep learning dus ook andere technieken, waaronder de genoemde traditionelere benaderingen die veel overheidsinstellingen gebruiken. Kortom, deze definitie is begrensd genoeg om AI te onderscheiden van algoritmes en digitale technologie in het algemeen, en open genoeg om toekomstige ontwikkelingen te omvatten. Figuur 1.1 geeft een overzicht van de besproken definities en de AI HLEG-definitie waar we in dit rapport van uitgaan.

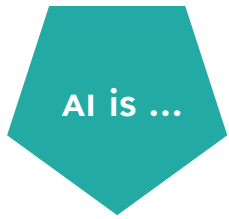
Het is goed om te benadrukken dat de huidige toepassingen die volgens deze definitie AI zijn, 'smalle' (*narrow AI*) of 'zwakke' AI (*weak AI*) betreffen.[20] De AI die wij nu kennen, is gericht op specifieke vaardigheden zoals beeld- of spraakherkenning en heeft weinig van doen met het volle spectrum van menselijke

19 Martin Ford (2018) interviewde 23 experts voor zijn boek *Architects of intelligence: The truth about AI from the people building it* en vroeg hen: "What year do you think human-level AI might be achieved, with a 50 procent probability?" De meesten wilden dat alleen anoniem doen en het gemiddelde waar ze mee kwamen was 2099, dus bijna 80 jaar van nu. In latere hoofdstukken komen we terug op de mogelijkheid van AGI.

20 Van de twee termen 'narrow AI' en 'weak AI' prefereren wij de eerste. De laatstgenoemde suggereert namelijk dat dit type niet sterk zou zijn, terwijl dat wel degelijk het geval kan zijn. Het is alleen beperkt tot een welomlijnd domein. Een computerprogramma kan bijvoorbeeld bijzonder sterk zijn in het vertalen van teksten, maar blijft smal, omdat het niet toegepast kan worden voor beeldherkenning.

cognitieve vermogens van AGI. Dat laat onverlet dat ook of juist huidige AI-toepassingen grote vraagstukken met zich meebrengen. De Amerikaanse hoogleraar Machine Learning Pedro Domingos heeft dit mooi verwoord. Wij richten ons volgens hem te veel op een toekomstige AGI en te weinig op de 'narrow AI' die al overal om ons heen is: "Mensen vrezen dat computers te slim worden en de wereld zullen overnemen, maar het echte probleem is dat ze te dom zijn en de wereld al hebben overgenomen."[21]

Figuur 1.1 Verschillende definities van AI



'Het gebruik van algoritmes. De term 'algoritme' verwijst naar een gespecificeerde instructie om een probleem op te lossen of om een berekening uit te voeren.'

'De nabootsing van alle menselijke intellectuele vermogens door computers.'

'De nabootsing van complexe menselijke vaardigheden door machines.'

'Technologie die adequaat kan functioneren door op een vooruitziende manier met de omgeving om te gaan.'

Het soort systemen dat intelligent gedrag vertoont door hun omgeving te analyseren en – met enige graad van autonomie – actie te ondernemen om specifieke doelen te bereiken.

Dat AI moeilijk te definiëren is, hangt dus samen met de evolutie van het vakgebied. Die evolutie gaan wij nu nader bekijken. Een historisch overzicht is niet alleen relevant als achtergrond om AI te kunnen begrijpen, maar is ook de opmaat om in het volgende hoofdstuk te zien dat AI momenteel op een keerpunt staat.

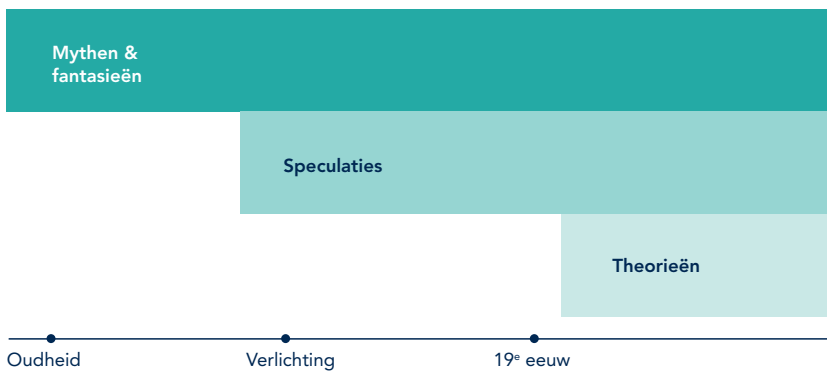
21 Domingos 2015: 286.

1.2 AI voorafgaand aan het lab

De geboorte van sommige disciplines is heel precies te dateren. AI is zo'n discipline. Het ontstaan ervan in het laboratorium wordt vaak gedateerd in 1956 tijdens een zomerschool aan Dartmouth College. AI is toen echter niet uit de lucht komen vallen. Voordat de technologie serieus in het lab onderzocht werd, had ze al een lange voorgeschiedenis.

Die voorgeschiedenis kent grofweg drie fasen: vroege mythische voorstellingen over artificiële vormen van leven en intelligentie, speculaties over denkende machines tijdens de Verlichting, en het leggen van het theoretische fundament voor de computer (zie figuur 1.2). Dit laatste vormde de springplank voor de ontwikkeling van AI als aparte discipline. We gaan hier in op deze drie fasen die voorafgingen aan het laboratoriumonderzoek naar AI. Deze fasen hebben elkaar overigens nooit uitgesloten. Parallel aan het theoretische onderzoek naar AI bleven er altijd mythen bestaan en werd er creatief gespeculeerd over de toekomst. De verschillende fasen geven aan hoe de aard en het accent in het denken over AI door de tijd is veranderd.

Figuur 1.2 Drie fases van AI voorafgaand aan het lab



De mythische voorstelling van AI

Al eeuwen bestaan er mythen en verhalen die gaan over wat we nu AI zouden noemen (zie figuur 1.3). Vooral de oude Grieken hadden in hun mythologie een veelheid aan figuren die als kunstmatige vormen van intelligentie zijn te karakteriseren.[22] Neem Talos, een robot gemaakt door de grote uitvinder Daedalus om het eiland Kreta te beschermen. Elke dag rende Talos rondjes om

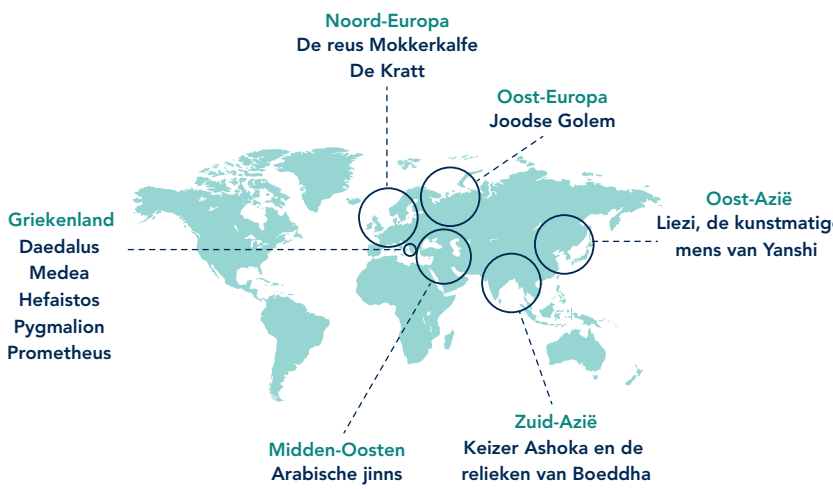
22 In *Gods and Robots – Myths, Machines, and Ancient Dreams of Technology* brengt Mayor (2018) het fenomeen '*made, not born*' in de Oudheid in kaart.

het eiland en als hij schepen zag naderen, dan gooide hij er stenen naartoe. Een mythe over een mechanische supersoldaat dus. Een robotisch exoskelet van het Amerikaanse leger draagt dan ook dezelfde naam.

Daedalus was de grote mythische uitvinder van de Oudheid. Hij is beroemd om de vleugels die zijn zoon Icarus het leven kostten, maar is tevens de uitvinder van allerlei kunstmatige intelligenties zoals bewegende standbeelden en dus de robot Talos. Volgens de mythe werd deze robot verslagen door de heks Medea, die hem misleidde om zichzelf onklaar te maken. Terwijl Daedalus dus een antieke AI-uitvinder was, heeft Medea in de mythen op magische wijze contact met AI. Haar vader was bovendien verantwoordelijk voor het ontstaan van kunstmatige soldaten die konden vechten zonder te stoppen.

Naast deze twee menselijke figuren van Daedalus en Medea, werden ook Griekse goden geassocieerd met kunstmatige intelligentie. Hefaistos, de smid van de goden, zou in zijn werkplaats bijgestaan worden door mechanische helpers. Hij bouwde ook gereedschap dat zichzelf bewoog en een hemelpoort die automatisch openging. De titaan Prometheus 'bouwde' de mens en stal voor hen het vuur van de goden. Als straf voor de mensheid creëerde Zeus een soort robot, de mechanische vrouw Pandora, die met haar vaas allerlei leed over de mensen uitstortte. Een minder grimmig voorbeeld is de mythe van Pygmalion. Deze beeldhouwer werd verliefd op een standbeeld dat hij gemaakt had, waarna Aphrodite haar tot leven wekte en zijn creatie met de naam Galatea zijn vrouw werd. In mythen maakten de oude Grieken dus al voorstellingen van wat wij nu *killer robots*, mechanische assistenten en seksrobots noemen.

Figuur 1.3 Oude mythes over AI

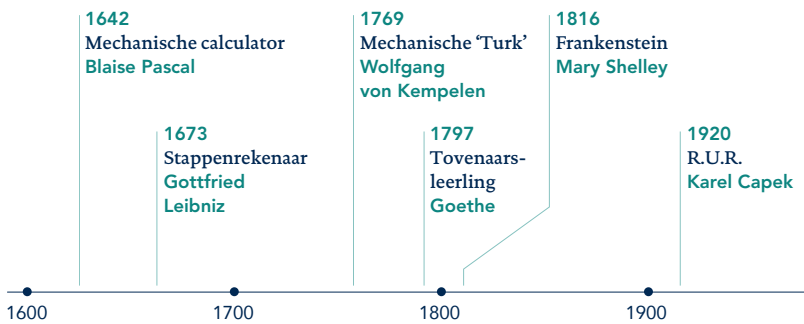


Ook in andere tradities bestaan verhalen over vormen van AI. Denk aan de joodse Golem en de Arabische mythen over 'jinns' die wensen in vervulling kunnen brengen. Het Boeddhistische verhaal *Lokapannatti* gaat over hoe de keizer Ashoka de relieken van de Boeddha in handen wilde krijgen, die echter door gevaarlijke mechanische bewakers – gemaakt in Rome – werden beschermd.[23] De Noorse mythologie kent de reus Mokkerkalfe, die gebouwd was om met Thor te strijden. De oude Chinese tekst *Liezi* verhaalt over de ambachtsman Yanshi die een kunstmatig mens bouwde met leer als spieren en hout als botten.[24] Uit Estland komt een mythe over de Kratt, een magisch wezen gemaakt van hooi en huishoudelijke artikelen die alles deed wat de eigenaar vroeg. Als de Kratt niet bezig werd gehouden, werd het een gevaar voor de eigenaar. De wet die in Estland werd afgekondigd en toeziet op de aansprakelijkheid voor het gebruik van algoritmes, wordt aldaar ook de 'Krattwet' genoemd.

Speculatie over de denkende machine

Een volgende fase brak aan als gevolg van de 'mechanisering van het wereldbeeld'[25] door het werk van mensen als Galileo Galilei, Isaac Newton en René Descartes. Samen met het mechanische wereldbeeld werden ook allerlei nieuwe machines gebouwd. Kunstmatige intelligentie lag nog ver buiten het bereik van de mogelijkheden, maar de nieuwe machines leidden wel tot speculatie over het ontstaan ervan (zie figuur 1.4). Die speculatie over AI was ditmaal niet langer mythisch maar mechanisch van aard.

Figuur 1.4 Tijdlijn van speculaties over AI



23 Zarkadakis 2015: 34.
24 Brynjolfsson en McAfee 2014: 250.
25 Beschreven door Dijksterhuis in *De mechanisering van het wereldbeeld* (1950).

Blaise Pascal bouwde in 1642 een mechanische calculator die volgens hem "dichter bij gedachten was dan alle handelingen van dieren".[26] In 1673 bouwde Gottfried Leibniz een apparaat genaamd 'de stappenrekenaar' (*step reckoner*), waarmee berekeningen gedaan konden worden. Daarmee werd de basis gelegd voor veel toekomstige rekenmachines.[27] Iets waar filosofen in die tijd over speculeerden onder de term 'automata'.

In 1769 bouwde Wolfgang von Kempelen een bijzondere geavanceerde machine – zo geloofde men in ieder geval voor lange tijd. Zijn mechanische 'Turk' bood hij aan de Oostenrijkse keizerin Maria Theresa aan, waarna hij wereldwijde faam verkreeg. Het grote apparaat was namelijk een automatische schaakmachine. Het apparaat toerde 48 jaar lang door de westerse wereld en versloeg mensen als Napoleon Bonaparte en Benjamin Franklin. Pas in de jaren twintig van de negentiende eeuw kwam men erachter dat het een grote zwendel was en dat er een mens in de machine zat die de stukken bewoog.[28] Het bedrijf Amazon heeft overigens een platform dat '*mechanical Turk*' heet. Daarop kunnen mensen goedkoop online klussen laten uitvoeren. Minder verhuld dan in Von Kempelens versie wordt het werk ook hier gedaan door mensen achter de schermen.

Speculatie over AI in deze tijd kon ook magische vormen aannemen. Goethes verhaal over de tovenaarsleerling, beroemd gemaakt door Disney's tekenfilm *Fantasia* met Mickey Mouse, gaat over een leerling die een spreuk gebruikt om een bezem water te laten halen. Hij kent de spreuk echter niet om het proces weer te stoppen en in plaats daarvan vermenigvuldigt de bezem zich en ontstaat er een fiasco totdat de tovenaar weer thuiskomt.[29] Andere magische verhalen over iets dat lijkt op AI, zijn Pinokkio en het horrorverhaal van W.W. Jacobs over een apenpoot die drie wensen op verschrikkelijke wijze in vervulling laat gaan.

Verhalen over magie liepen ook over in verhalen die een beetje dichter bij de wetenschap stonden, namelijk in het genre van sciencefiction. In 1816 kwam een groep schrijvers in Genève bijeen die lange tijd binnenshuis moesten blijven vanwege een vulkaanuitbarsting in Indonesië. Dat was het 'jaar zonder zomer' waarin regen de mensen binnen hield. Geïnspireerd op de magische verhalen van E.T.A. Hoffman stelde Lord Byron voor dat alle schrijvers een bovennatuurlijk verhaal zouden schrijven. Mary Shelley maakte daarvoor de eerste versie van haar beroemde verhaal over Frankenstein.[30] Dit verhaal van een

26 Russell 2019: 40.
27 Broussard 2019: 76.
28 Zarkadakis 2015: 37.
29 Wiener 1964: 57.
30 Zarkadakis 2015: 60-63.

wetenschapper die een kunstmatige vorm van leven creëert die zich uiteindelijk tegen zijn maker keert, is het archetypische verhaal geworden van het gevaar van de moderne techniek. Dit motief leeft voort in talloze films, waarvan *Blade Runner* (1982), *The Terminator* (1984) en *The Matrix* (1999) inmiddels klassiekers zijn.

Een ander belangrijk sciencefictionwerk in het kader van speculatie over AI is het boek R.U.R. van de Tsjechische schrijver Karel Capek. In dit boek introduceert de schrijver de term 'robot', die is afgeleid van het Oudkerkslavische woord 'rabota', dat herendienst of dwangarbeid betekende. Ook uit dit verhaal spreekt een klassieke angst voor AI. In het verhaal worden in een fabriek kunstmatige arbeiders gemaakt, de 'roboti', die in opstand komen tegen hun makers en uiteindelijk de mensheid vernietigen.[31] Het boek van Capek verscheen in 1920, toen inmiddels allang een volgende fase was ingezet die het denken over AI een stuk concreter zou maken.

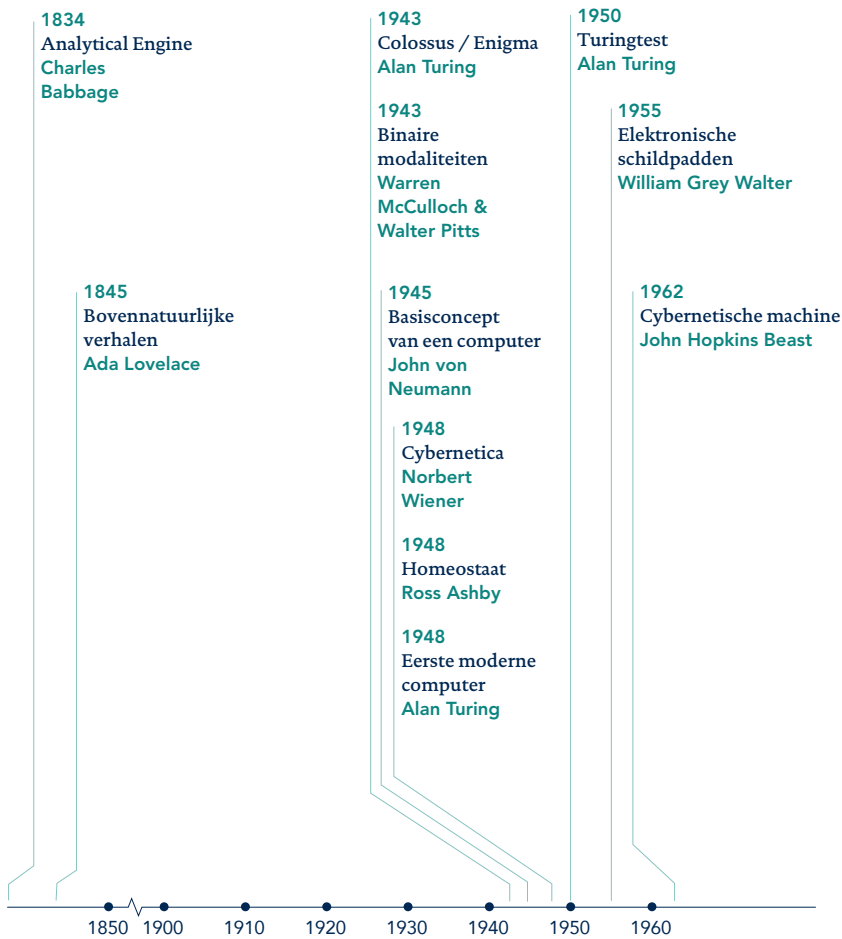
De theorie van AI

Vanaf de tweede helft van de negentiende eeuw wordt het denken over AI als 'denkende computers' minder fantastisch en neemt het serieuze theoretische vormen aan (zie figuur 1.5). Dat moment loopt dan ook parallel met de theorie en bouw van de eerste computers.

Ada Lovelace – de dochter van de dichter Byron die in Genève met vrienden bovennatuurlijke verhalen schreef – heeft daar in de jaren veertig van de negentiende eeuw een belangrijke rol in gespeeld. Lovelace voorzag een machine die op basis van logica zelf complexe muziekstukken kon maken en ook wetenschappelijk onderzoek verder zou brengen. Haar kennis Charles Babbage ontwierp in 1834 een dergelijk apparaat, de '*Analytical Engine*' genaamd.[32] Toen hij er niet in was geslaagd om zijn eerdere enorm complexe '*Difference Engine*' te bouwen, maakte hij het ontwerp voor de *Analytical Engine*, waarmee hij wiskundige en astronomische tabellen wilde bouwen.[33] Lovelace zag echter een veel breder gebruik voor een 'denkende machine' die kon redeneren over "alle onderwerpen in het universum".[34] Lovelace schreef zelf ook programma's voor de hypothetische machine. De wetenschap was in die tijd echter nog niet ver genoeg om daadwerkelijk computers te bouwen.

31 Rid 2016: 83.
32 Boden 2018: 6.
33 Freeman en Louçã 2001: 309.
34 Russell 2019: 40.

Figuur 1.5 Tijdlijn van theorieën over AI



Dat gebeurde pas tijdens de Tweede Wereldoorlog. Rekenkracht was nodig voor de verdediging tegen de snelle bombardementen van de Duitsers vanuit de lucht. Het gebruik van vliegtuigen om bommen af te vuren maakte het onmogelijk voor het menselijk oog om bijtijds te reageren met afweersystemen. Daartoe moesten de trajecten van snel bewegende objecten berekend kunnen worden. Het onderzoek hiernaar legde de basis voor de moderne computer en voor een discipline die vanaf de jaren vijftig op zou komen, die van de cybernetica. Rondom dit onderzoek kwamen meteen al vragen over automatisering en menselijke controle op die nog steeds actueel zijn: "De tijdsfactor is zo klein geworden voor alle operatoren dat de menselijke link, die de enige onveranderlijke factor in het hele probleem lijkt en die bijzonder wispelturig is, meer en

meer de zwakste schakel in de keten van operaties is geworden totdat het duidelijk wordt dat de menselijke link uit de sequentie verwijderd moet worden", aldus een defensiewoordvoerder destijds.[35]

Tijdens de Tweede Wereldoorlog kreeg de ontwikkeling van de computer ook een impuls door het Britse onderzoeksprogramma *Colossus* om Enigma, het geheime communicatiesysteem van de Nazi's, te kraken. Een van de mensen die daar in het grootste geheim op Bletchley Park aan werkte, was Alan Turing, veelal beschouwd als de vader van zowel de computer als AI. Hij werkte aan de bouw van de eerste echte moderne computer in Manchester in 1948. In 1950 schreef hij een paper waarin hij een gedachtenexperiment voorstelde voor een '*imitation game*' waarbij een computer zich voordoet als mens.[36] Dat spel is bekend komen te staan als de Turingtest. Een computer slaagt voor die test als een mens niet kan vaststellen of de geschreven antwoorden op zijn vragen afkomstig zijn van een mens of een computer. Varianten van deze test worden nog steeds gebruikt wanneer bijvoorbeeld AI-systemen vergeleken worden met menselijke vermogens om beelden te herkennen of taal te gebruiken.[37]

Een andere belangrijke theoretische bijdrage aan het vakgebied was een paper van de psychiater en neuroloog Warren McCulloch en de wiskundige Walter Pitts.[38] In dat paper verenigden zij het werk van Turing over computers met de propositielogica van Bertrand Russell en de theorie van neurale synapsen van Charles Sherrington. De kern van deze bijdrage was dat zij op verschillende domeinen binaire modaliteiten (een situatie met twee opties) aantoonden en daarmee een gemeenschappelijke taal voor neurofysiologie, logica en computatie ontwikkelden. Het onderscheid tussen 'waar en onwaar' in de logica werd zo verbonden met het 'aan of uit' staan van neuronen en de computerwaarden van '0 en 1' in Turingmachines.[39]

---

35 Rid 2016: 37-38.

36 Turing 2009 [1950].

37 Er is ook kritiek op het gebruik van taal in de Turingtest. Yann LeCun, prominent AI-wetenschapper, stelt in een interview dat er vormen van intelligentie bestaan die niets met taal te maken hebben (Ford 2018: 129). Zo zijn er dieren met een minder complexe taal dan mensen, die desalniettemin goede modellen van de wereld hebben en gereedschap kunnen gebruiken.

38 McCulloch en Pitts 1943.

39 In een lezing aan Yale in de jaren vijftig verwoordde de wetenschapper John von Neumann de overeenkomst tussen de computer en het brein als volgt: "The nervous pulses can clearly be viewed as (two-valued) markers, in the sense discussed previously: the absence of a pulse represents one value (say, the binary digit 0), and the presence of one represents the other (say, the binary digit 1)." Von Neumann 2012 [1958]: 43.

John von Neumann ontwikkelde het basisconcept van een computer verder met componenten als de centrale processor, geheugen en input-outputapparatuur.[40] Eveneens een belangrijke grondlegger van de theorie voor AI was Norbert Wiener. Hij muntte de term 'cybernetica' in 1948 voor "de studie naar controle en communicatie in dier en machine".[41] Het kernidee daarvan is dat mensen, dieren en machines volgens een aantal basisprincipes te begrijpen zijn. Het eerste principe is controle. Al deze entiteiten streven ernaar entropie tegen te gaan en hun omgeving te controleren. Dit doen zij middels het principe van feedback of terugkoppeling. Dat is het "vermogen om toekomstig gedrag aan te passen aan de ervaring uit het verleden". Middels continue aanpassing en terugkoppeling zorgen organismen en machines ervoor dat een equilibrium, of homeostase, wordt gerealiseerd. Thermostaten en servomechanismen waren centrale metaforen voor Wiener om deze processen toe te lichten. Alhoewel cybernetica als aparte wetenschappelijke discipline niet lang bleef bestaan, werken de kernconcepten ervan door in allerlei disciplines.[42]

Tekstbox 1.1 – De homeostaat en elektronische schildpadden

In 1948 onthulde de Brit Ross Ashby zijn uitvinding van de 'homeostaat', een machine die vier elektromagneten in een stabiele positie wist te houden. Over dit 'protobrein' schreef de *Herald* dat jaar: "Het klikkende brein is slimmer dan dat van de mens."[43] Een ander hoogtepunt van de cyberneticabeweging waren de elektronische schildpadden van William Grey Walter in de jaren vijftig. Deze kleine apparaten konden rondlopen zonder in problemen te komen en hun oplader ergens in de ruimte om hen heen vinden als hun batterij zwak was. Als groep vertoonden ze bovendien complex sociaal gedrag. Een later voorbeeld van een cybernetische machine was de John Hopkins Beast, die begin jaren zestig door gangen bewoog om met behulp van sonar en een fotoceloog op zoek te gaan naar een oplaadpunt.[44]

40 Freeman en Louçã 2001: 310.
41 Wiener 2019 [1965].
42 Rid 2016: 47-52. Beroemde cybernetici in verschillende disciplines waren onder andere de neurofysioloog Warren McCulloch, de fysicus Heinz von Foerster, de managementtheoreticus Stafford Beer, de filosoof Humberto Maturana, de politicoloog Karl Deutsch, de antropoloog Gregory Bateson en de socioloog Talcott Parsons.
43 Rid 2016: 53-55.
44 Moravec 1988: 7.

Met de ontwikkelingen in deze periode werd de wetenschap rijp om het dromen van en denken over AI om te zetten in het daadwerkelijk ontwikkelen van de technologie en het ermee experimenteren in het laboratorium. Het startschot daarvoor vond zoals gezegd plaats in 1956.

Kernpunten – AI voorafgaand aan het lab

- Al eeuwen geleden waren er mythische voorstellingen van AI. De meest prominente voorbeelden daarvan zijn de verhalen van de oude Grieken over Daedalus, Medea, Hefaistos, Prometheus en Pygmalion.
- De mechanisering van het wereldbeeld vanaf de zeventiende eeuw maakte de bouw van allerlei machines mogelijk. Daarmee gepaard gingen ook speculaties over mechanische breinen.
- Vanaf de Industriële Revolutie wordt er fictie geschreven over kunstmatige intelligentie, zoals Frankenstein en R.U.R.
- De theoretische fundamenten voor AI worden gelegd in samenspel met de bouw van de eerste computers door mensen als Alan Turing.

1.3 AI in het lab

De eerste golf

Het begin van AI als discipline kan, zoals gezegd, heel precies gedateerd worden.[45] Na mythen, speculaties en theoretisering belandde AI in 1956 in het lab toen een groep wetenschappers het als onderwerp vaststelde voor een specifiek evenement: het *Dartmouth Summer Research Project on Artificial Intelligence*. Dat was een brainstormbijeenkomst die zes weken duurde en waar verschillende grondleggers van de discipline bij aanwezig waren. De organisatoren waren bijzonder optimistisch over wat zij met deze groep in een aantal weken zouden kunnen bereiken. Dat blijkt uit het voorstel dat zij aan de Rockefeller Foundation schreven:

> "Wij stellen een studie voor van twee maanden van tien man naar kunstmatige intelligentie. (…) De studie zal plaatsvinden op basis van de vooronderstelling dat elk aspect van leren of elk aspect van intelligentie in principe zo precies beschreven kan worden dat een machine het kan simuleren. Een poging zal worden gedaan om erachter te

45 De geschiedenis van een vakgebied kan op verschillende manieren worden geschreven. Nadruk kan liggen op de fundamentele wetenschap of op uitvindingen en toepassingen. Denk aan het verschil tussen de ontwikkeling van de natuurwetenschap en de Industriële Revolutie. In dit hoofdstuk combineren wij beide perspectieven, maar het idee van golven in AI volgt vooral uit het perspectief van uitvindingen en toepassingen.

> komen hoe machines taal kunnen gebruiken, abstracties en concepten te laten ontwikkelen, problemen op te lossen die nu voor mensen zijn gereserveerd en om zichzelf te verbeteren. Wij denken dat significante vooruitgang op een of meer van deze problemen geboekt kan worden als een nauwkeurig geselecteerde groep wetenschappers er een zomer lang aan samenwerkt."[46]

Het voorstel was overambitieus en op alle genoemde gebieden wordt ook in de huidige tijd nog volop onderzoek gedaan. Maar met dit project formuleerden deze wetenschappers wel een onderzoeksagenda waarmee zij de discipline van AI lanceerden.

Het zomerproject werd georganiseerd door John McCarthy en Marvin Minsky. Het was McCarthy die in 1956 de term '*Artificial Intelligence*' bedacht. Marvin Minsky was een leidend figuur in de geschiedenis van AI en is door de jaren heen wereldwijd aan allerlei prominente hightechprojecten verbonden geweest. Samen zetten McCarthy en Minsky ook het Artificial Intelligence Lab aan MIT op, later omgedoopt tot het MIT Media Lab en nog steeds een centrum voor het creatief gebruik van nieuwe technologie.[47] Tot de aanwezigen behoorden ook Herbert Simon, Nobelprijswinnaar in de economie en winnaar van de Turing Award, onder andere bekend van het idee van '*bounded rationality*' alsook oprichter van het Carnegie Institute of Technology, John Nash, wiskundige, speltheoreticus en eveneens Nobelprijswinnaar in de economie, en Arthur Samuel, pionier in computerspellen en degene die de term 'machine learning' populariseerde. Deze topwetenschappers brachten AI naar het lab.

Jaren volgden van groots optimisme en brede interesse in het veld. Die periode is bekend komen te staan als de eerste AI-lente of -golf. Er werden verschillende programma's ontwikkeld voor het bordspel dammen, hoewel deze in die periode zeker nog niet erg goed waren. Het programma ontwikkeld door de genoemde Arthur Samuel slaagde er uiteindelijk in om diens menselijke maker te verslaan en dat deed veel stof opwaaien, al stond Samuel niet per se bekend als een groot damspeler. Wiener schreef bijvoorbeeld in 1964 dat alhoewel Samuel na wat instructie het programma uiteindelijk weer kon verslaan, "de methode van leren in principe niet anders is dan de methode waarmee mensen leren dammen". Hij verwachtte bovendien dat hetzelfde met schaken zou gebeuren in

46 Bostrom 2016: 6.
47 Broussard 2019: 69-70.

tien tot vijfentwintig jaar en dat mensen ondertussen hun interesse voor beide spellen zouden verliezen.[48]

Spannende doorbraken kwamen vervolgens van AI-systemen die zich richtten op een ander soort vraagstukken: logische en conceptuele vraagstukken. Zo was de *Logic Theory Machine* gebouwd om logische theorema's te bewijzen. Het systeem slaagde er niet alleen in om bewijs te leveren voor achttien van Bertrand Russells logische theorema's, het ontwikkelde ook nog eens een eleganter bewijs voor één daarvan. Dat was belangrijk, want terwijl Samuel een middelmatig damspeler was, was Bertrand Russell een vooraanstaand logicus.

Een volgende mijlpaal was de *General Problem Solver*. Dat was een programma dat in principe op elk probleem toegepast moest kunnen worden, vandaar de naam. Door problemen te vertalen in doelen, subdoelen, acties en operatoren, kon dit programma vervolgens beredeneren wat het juiste antwoord was. Een voorbeeld van een probleem dat *General Problem Solver* oploste, is de klassieke logische puzzel van de rivieroversteek.[49]

Halverwege de jaren zestig waren de eerste studenten van de AI-pioniers bezig met programma's die geometrische theorema's bewezen, intelligentietesten oplosten evenals algebrateksten en calculusexamens. De discipline maakte dus voortgang, maar de effecten ervan buiten het lab waren zeer beperkt. Er waren interessante experimenten met robots, zoals eind jaren zestig op het Stanford Research Institute met *Shakey the Robot*, die zijn weg vond op basis van logische redeneringen.[50] Het Amerikaanse technologiebedrijf General Electric maakte imposante robots zoals de *Beetle* en een exoskelet waarmee mensen zware gewichten konden tillen.[51] Erg praktisch waren deze robots overigens niet.

Tegelijkertijd waren de verwachtingen hooggespannen. In 1965 voorspelde Herbert Simon dat "machines binnen twintig jaar in staat zullen zijn om al het werk te doen dat mensen kunnen doen".[52] En de Britse wiskundige Irving Jack Good voorzag een door machines veroorzaakte "*intelligence explosion*". Dat zou dan tegelijkertijd de laatste uitvinding van de mensheid zijn, omdat machines

48 Wiener 1964: 22-24. Het zou uiteindelijk dertig jaar duren voordat een schaakgrootmeester verslagen werd, zoals we straks zullen zien. Interesse in de spellen is in ieder geval niet verdwenen sinds er geavanceerde programma's voor zijn.

49 Boden 2018: 10. Bij dit logische probleem moeten drie entiteiten allemaal de rivier oversteken. Per oversteek kunnen er twee entiteiten mee. De ene entiteit bedreigt de andere en dus kan niet elk duo samen de oversteek maken. Het is de vraag met welke combinaties het mogelijk is om iedereen ongeschonden naar de overkant te krijgen.

50 Russell 2019: 52.

51 Rid 2016: 136.

52 Brynjolfsson en McAfee 2014: 141.

vervolgens de intelligentste wezens op aarde zouden zijn en dus alle verdere uitvindingen zouden doen.[53]

AI sprak ook tot de verbeelding van mensen buiten de wetenschap. In 1967 werd het computerprogramma MacHack VI tot erelid van de Amerikaanse schaakfederatie gemaakt, ondanks dat het nog maar weinig wedstrijden had gewonnen.[54] Een paar jaar later verscheen de film *Colossus: The Forbin Project*. Daarin wordt de controle over het Amerikaanse militaire arsenaal aan een computerprogramma gegeven, omdat het beslissingen kan nemen die superieur zijn aan die van mensen en niet gehinderd wordt door emoties. Wanneer de Sovjets vervolgens een vergelijkbaar programma onthullen, gaan de twee programma's op voor mensen onbegrijpelijke manier met elkaar communiceren en nemen ze vervolgens het bestuur van de hele wereld in handen. Hun geprogrammeerde doel van vrede wordt bereikt, maar tegen de prijs van menselijke vrijheid.

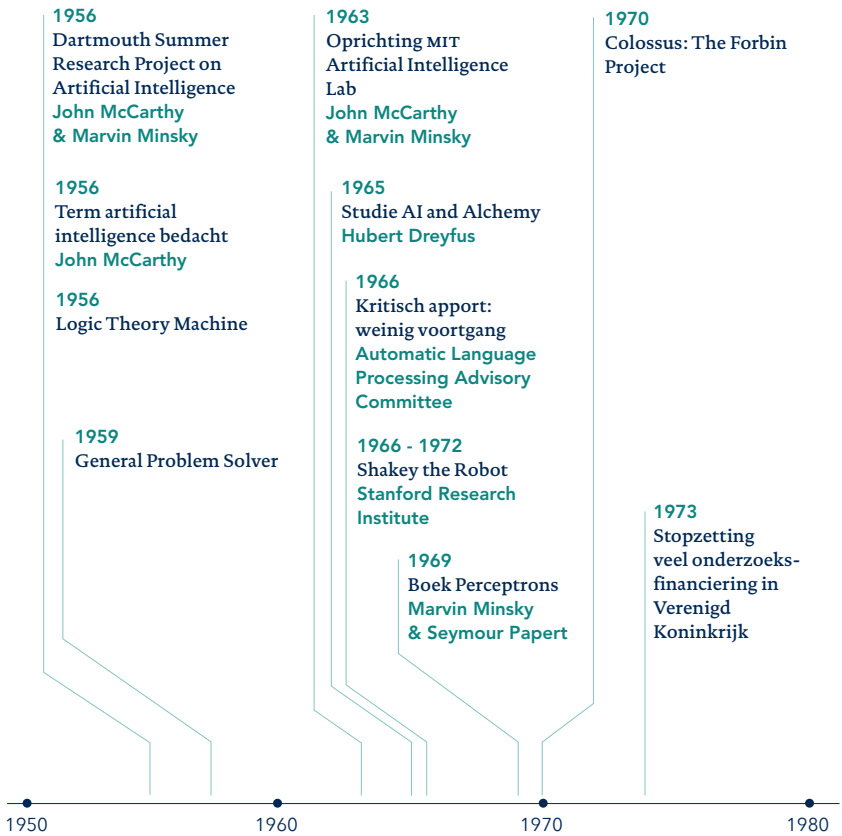
De kloof tussen verwachting en werkelijkheid bleef niet onopgemerkt en vanaf de tweede helft van de jaren zestig nam de kritiek op AI-onderzoek toe. De filosoof Hubert Dreyfus zou zijn leven lang kritisch blijven over de mogelijkheden van AI. In 1965 schreef hij in opdracht van de RAND Corporation – het onderzoeksinstituut van het Amerikaanse leger – de studie *AI and Alchemy*, waarin hij de komst van intelligente machines in de nabije toekomst onmogelijk achtte. In 1966 rapporteerde de Automatic Language Processing Advisory Committee aan de Amerikaanse regering dat er maar weinig voortgang was, waarna onderzoeksgeld van de National Research Council opdroogde. In 1973 hield Sir James Lighthill op verzoek een enquête waaruit kritiek naar boven kwam over het falen van AI om de grandioze doelen te bereiken die in het vooruitzicht waren gesteld. Hierdoor werd ook in het Verenigd Koninkrijk veel onderzoeksfinanciering stopgezet.[55]

Een probleem waar veel AI-systemen destijds tegenaan liepen was de zogenoemde 'combinatorische explosie'. Deze systemen losten problemen op door alle mogelijke opties te onderzoeken. Die methode liep bij vraagstukken met heel veel mogelijke combinaties al gauw tegen de grenzen van de rekenkracht aan. Meer heuristische benaderingen waren nodig, die aan de hand van vuistregels het aantal combinaties kunnen reduceren. Die waren toen echter nog niet voorhanden. Samen met andere problemen, zoals het gebrek aan data om de systemen te voeden en beperkte capaciteit van de hardware, stokte de

53 Rid 2016: 148. Later zou de schrijver Vernor Vinge de term 'singulariteit' (*singularity*) voor dit scenario munten.
54 Bakker en Korsten 2021: 24.
55 Leung 2019: 253.

vooruitgang. Toepassingen van AI in de praktijk bleken onbetrouwbaar. In de jaren zestig werd AI bijvoorbeeld ontwikkeld om tijdens de Koude Oorlog Russische communicatie te vertalen. Dat leverde weinig succesvolle resultaten op. Een beroemd voorbeeld is de vertaling van "de geest ('spirit') is welwillend, maar het vlees ('flesh') zwak" als "de vodka is goed, maar het vlees ('meat') is verrot".[56] Het optimisme sloeg in de loop van de jaren zeventig daarom om. Er waren te weinig doorbraken, de kritiek groeide en de financiering droogde op. De eerste 'AI-winter' was aangebroken en maakte een einde aan de eerste AI-golf. Figuur 1.6 geeft een overzicht van de opkomst van AI als discipline.

Figuur 1.6 Tijdlijn van de opkomst van AI als discipline (eerste golf)



56 Russell en Norvig 2021: 21.

Twee benaderingen

Belangrijk is hier te vermelden dat tijdens de eerste golf twee prominente benaderingen binnen het veld van AI ontstonden. Weliswaar zijn er ook nog andere benaderingen die wij later zullen noemen, maar deze twee benaderingen, die al in de eerste golf tot ontwikkeling kwamen, zijn tot op heden dominant in het veld. De eerste is gebaseerd op regels (*rule-based*), maar wordt ook wel aangeduid met namen als 'symbolische' of 'logische' AI. Vanaf de jaren zeventig zijn binnen deze benadering de zogenoemde 'expertsystemen' opgekomen. De kern van deze benadering is dat computers leren door logische regels te coderen met formules van het type "als X, dan Y". Het gebruik van logica en regels is waarom ook wel gesproken wordt van symbolische AI, omdat deze benadering regels volgt die uit te drukken zijn in menselijke symbolen.

De tweede benadering maakt gebruik van artificiële neurale netwerken (ANNs). Er wordt ook wel gesproken van *connectionisme*. Binnen deze benadering vallen *deep learning* en *parallel distributed processing*, waar de laatste jaren veel aandacht voor is. Het centrale idee bij deze benadering is om de werking van neuronen in het menselijk brein te simuleren. Daarvoor worden sets gebouwd van een soort artificiële neuronen in netwerken die informatie kunnen ontvangen en versturen. Die netwerken worden vervolgens met een grote hoeveelheid data gevoed waaruit zij zelf patronen moeten destilleren. Bij deze benadering stellen mensen dus niet vooraf regels op. De meeste ANNs zijn gebaseerd op het principe dat Donald Hebb, een Canadese psycholoog, al in 1949 in zijn boek *The Organization of Behavior* formuleerde: "*neurons that fire together, wire together*"[57] – als twee neuronen vaak samen worden geactiveerd, worden ze aan elkaar verbonden.

Beide stromingen waren er al vanaf het begin. Terwijl veel van de grondleggers uit 1956 de regelgebaseerde benadering volgden, werd rond die tijd ook het eerste artificiële neuron aan Cornell University gemaakt.[58] Het verschil tussen de twee benaderingen kan als volgt worden uitgelegd. Om een kat op een foto te herkennen, wordt in de eerste benadering een reeks van 'als-dan'-regels opgesteld over hoe een kat te herkennen: als bepaalde kleuren aanwezig zijn, een aantal ledematen, bepaalde vormen in het gezicht, snorharen te zien zijn, enzovoort, dan is er sprake van een kat. Met die regels 'redeneert' een programma over de data.

In de tweede benadering zou het programma bijvoorbeeld een groot aantal foto's te zien krijgen die gelabeld zijn als 'kat' en 'niet-kat'. Op basis daarvan gaat het programma zelf patronen destilleren, waardoor het vervolgens op nieuwe

57 Domingos 2015: 93.
58 Greenfield 2017: 214.

foto's de aanwezigheid van een kat moet herkennen. Een andere variant binnen deze benadering maakt geen gebruik van labels maar toont grote aantallen afbeeldingen om het programma vervolgens zelf tot een clustering van katten te laten komen. In beide varianten zijn het echter niet door mensen geprogrammeerde regels, maar door het programma ontdekte patronen die leidend zijn.

Tijdens de eerste AI-golf werden beide sporen, zoals gezegd, verkend. Een voorbeeld van een toepassing van neurale netwerken waren Frank Rosenblatt's 'perceptrons', die letters leerden herkennen zonder dat die voorgeprogrammeerd waren. Dit sprak in de media van de jaren zestig bijzonder tot de verbeelding. Symbolische AI bleef echter dominant. De eerdergenoemde *Logical Theory Machine* en de *General Problem Solver* zijn daar voorbeelden van. Decennialang zou dit de belangrijkste benadering binnen AI blijven.

Vanuit symbolische AI was er ook veel kritiek op neurale netwerken. Door de afwezigheid van regels zou deze benadering onbetrouwbaar en maar beperkt toepasbaar zijn. Marvin Minsky, een fervent aanhanger van de symbolische benadering, schreef in 1969 samen met Seymour Papert het boek *Perceptrons*. Dat was een grondige kritiek van de benadering van neurale netwerken met wiskundige bewijzen van problemen die het niet kon oplossen. Voor veel mensen leek dat de doodsteek voor die benadering.[59] Dergelijke kritiek marginaliseerde niet alleen de positie van neurale netwerken, maar droeg ook bij aan de eerste AI-winter.

De tweede golf

In 1982 verkoos *Time Magazine* de *personal computer* tot 'Man of the Year'. Op datzelfde moment kwam er een opleving in AI en ontlook een tweede lente voor het vakgebied. De programmeertaal Prolog werd toentertijd gebruikt voor veel logische redeneersystemen. In 1982 deed de Japanse regering een immense investering in op Prolog gebaseerde AI met het *Fifth-Generation Computer Systems Project*.[60] Dat was een grootschalige tienjarige samenwerking tussen de overheid en het bedrijfsleven. Door een 'parallelle *computing* architectuur' op te zetten wilden de Japanners een impuls aan het veld geven. Het was de tijd van angst voor de Japanse economische groei en verschillende westerse landen volgden snel met hun eigen projecten.

De VS ontwikkelde de *Microelectronics and Computer Technology Corporation* (MCC) als een onderzoeksconsortium om competitief te blijven. Daar startte Doug Lenat in 1984 het enorme project CYC, waar Marvin Minsky een groot aanhanger van was. In het programma worden grote hoeveelheden menselijke

59 Uit een interview met Geoffrey Hinton (Ford 2018: 83).
60 Russell 2019: 271.

kennis ingevoerd.[61] In 1983 kondigde DARPA, de wetenschappelijke tak van het Amerikaanse leger, het *Strategic Computing Initiative* (SCI) aan om gedurende tien jaar een miljard dollar in het veld te investeren.[62] Zowel in het Japanse als in het Amerikaanse onderzoek werd AI breed benaderd, waarbij bijvoorbeeld ook hardware en menselijke interfaces een belangrijke rol speelden.[63] In 1983 kondigde het Verenigd Koninkrijk haar reactie op de Japanse plannen aan, met het *Alvey Programme*.

Een belangrijke ontwikkeling tijdens deze tweede golf was de opkomst van expertsystemen binnen de symbolische AI in de jaren zeventig. Dit is een vorm van regelgebaseerde AI waarbij menselijke experts op een bepaald domein worden gevraagd de regels te formuleren voor een programma. Een voorbeeld was MYCIN, een programma dat getraind werd door medische experts om artsen te adviseren bij het identificeren van infectieziekten en het voorschrijven van medicatie. DENDRAL werd gebruikt bij de analyse van moleculen in de organische chemie. Expertsystemen werden ook ontwikkeld voor de planning van fabrieken en om complexe wiskundige vraagstukken op te lossen, zoals het programma MACSYMA.

In Nederland werden eveneens in de jaren tachtig expertsystemen ontwikkeld en in pilots getest, onder meer voor de uitvoering op het terrein van de sociale zekerheid en de straftoemeting.[64] Mede dankzij specifieke onderzoeksprogrammering en -financiering door niet alleen NWO en diverse universiteiten maar ook door ministeries, kon ons land zich zelfs internationaal profileren met een relatief grote onderzoeksgemeenschap op het terrein van de rechtsinformatica. De expertsystemen vonden dus praktische toepassingen buiten het lab. Het Amerikaanse Office for Technology Assessment noemde expertsystemen "de eerste echte commerciële producten van ongeveer vijfentwintig jaar AI-onderzoek".[65] In 1984 stond op de voorpagina van de *New York Times* dat expertsystemen het vooruitzicht bieden op "computerondersteunde beslissingen die gebaseerd zijn op meer wijsheid dan een enkel mens kan bevatten".[66]

Desondanks vielen de resultaten van deze tweede golf uiteindelijk tegen. De hoge ambities van de grote nationale projecten werden niet gehaald, noch in Japan noch in de VS of Europa. Door beperkte resultaten werd de financiering

61 Domingos 2015: 35.
62 Leung 2019: 254.
63 Russell en Norvig 2020: 24.
64 Hage en Verweij 1999.
65 Leung 2019: 259.
66 Dreyfus en Dreyfus 1986: ix.

van het Amerikaanse SCI drastisch gereduceerd. Het waren onder andere problemen in de hardware die de mogelijkheden van deze projecten beperkten. Verschillende gespecialiseerde bedrijven op dit gebied gingen eind jaren tachtig failliet.[67] Maar ook expertsystemen zelf kenden hun problemen. Die konden al snel heel complex worden, kleine fouten in de regels hadden desastreuze gevolgen voor de uitkomsten en systemen faalden als twee regels elkaar tegenspraken.[68] Het project CYC loopt nog altijd, maar heeft in bijna vier decennia de hoge verwachtingen niet waargemaakt.[69] Zo zette eind jaren tachtig opnieuw een winter in en strandde de tweede golf.

De derde golf

In de jaren negentig kwam het veld van AI weer prominent in de aandacht en bloeide het opnieuw op. De benadering van logische systemen boekte een aantal successen. Een van de meeste iconische daarvan was de overwinning van IBM's programma Deep Blue op de schaakgrootmeester Garry Kasparov in 1997. Toentertijd was de gedachte dat het om fundamentele doorbraken in AI ging. De opvolger van dit programma, genaamd Watson, deed later mee aan het televisieprogramma *Jeopardy!*, waarbij deelnemers vragen moesten formuleren bij antwoorden. In 2011 versloeg het programma de regerende menselijke kampioenen. Dit werd gezien als bewijs dat AI nu dichtbij grip op de menselijke taal was gekomen, een belangrijke doorbraak. Beide prominente evenementen zijn voorbeelden van het gebruik van symbolische AI waarbij de lessen van schaakmeesters en de antwoorden van eerdere *Jeopardy!*-spelers als regels aan de programma's werden meegegeven. Tegelijkertijd groeide onder experts op dat moment juist de onvrede met deze benadering.

Hoewel beide gebeurtenissen voor het publiek grote doorbraken leken, ligt de werkelijkheid genuanceerder. Stuart Russell beschrijft hoe Claude Shannon in 1950 de basis van schaakalgoritmes legde en hoe er vervolgens innovaties volgden in de jaren zestig. Daarna verbeterden deze programma's zich volgens een voorspelbaar patroon, parallel met de groei van de rekencapaciteit. Dit kon gemeten worden aan de hand van de score die schaakspelers hebben. Om uit te komen bij de score van een grootmeester kom je met een lineair patroon uit in de jaren negentig, precies toen Deep Blue Kasparov versloeg. Het was dus niet zozeer een doorbraak, maar een te verwachten mijlpaal in een voorspelbaar

---

67 Lueng 2019: 255.

68 De grenzen van menselijk gedrag en taal proberen te vatten in regels is al eerder verkend door filosofen als Ludwig Wittgenstein (Wittgenstein 1984).

69 Volgens Ray Kurzweil, een aanhanger van neurale netwerken, heeft CYC zelfs bijna niets opgeleverd (Ford 2018: 233). Dat is te simpel. Dergelijke projecten staan aan de basis van technieken als *knowledge graphs* die tegenwoordig belangrijk zijn voor zoekopdrachten in Google of navigatie. Dat toont ook aan dat de twee benaderingen elkaar niet uitsluiten en in de praktijk vaak samengaan.

patroon.[70] Met brute rekenkracht had Deep Blue gewonnen. Bovendien hadden allerlei schaakkampioenen heuristische principes aan de software van het programma toegevoegd. In plaats van de slimme computer die de mens versloeg, kon dit ook gezien worden als een collectief van een computerprogramma en vele menselijke spelers tegenover een enkele grootmeester.[71] Het waren dus mens en machine die *samen* superieur waren aan een menselijke tegenstander.

Ook bij de overwinning van de computer bij het spel *Jeopardy!* zijn vraagtekens te plaatsen. Het is niet correct om te stellen dat het programma de complexe natuurlijke taal van mensen wist te begrijpen. Het spel volgt namelijk een heel formeel principe van vraag en antwoord en veel van de vragen hebben het karakter van een typische Wikipediapagina. Daardoor zijn ze relatief gemakkelijk te beantwoorden voor een programma dat snel bergen informatie kan doorzoeken op signaalwoorden. Een echt begrip van de taal was hiervoor niet nodig.

Terwijl deze logische systemen vanaf de jaren negentig de aandacht trokken, werd elders in het AI-veld al een tijd vooruitgang geboekt, waardoor het momentum uiteindelijk richting de benadering van neurale netwerken zou verschuiven. Die verschuiving begon al halverwege de jaren tachtig, toen fundamenteel onderzoek naar het zogenoemde '*backpropagation algorithm*', waarbij meerdere lagen van neurale netwerken getraind worden, de patroonherkenning van deze benadering verbeterde. Rond die tijd erkende het Amerikaanse ministerie van Defensie ook dat zij met haar financiering de benadering van neurale netwerken onterecht verwaarloosd had. Onder de noemer van '*parallel distributed processing*' kwamen de neurale netwerken in 1986 weer terug op het toneel. Het jaar daarvoor had John Haugeland in een boek de term 'GOFAI' geïntroduceerd: *good old-fashioned AI*, dat sindsdien als een pejoratieve term wordt gebruikt voor symbolische AI. In die tijd was Judea Pearl bezig om in plaats van logische redeneringen de waarschijnlijkheidstheorie in AI toe te passen.

Onder de radar vonden dus doorbraken plaats die tegen de dominante regelgebaseerde stroming ingingen. Een paper over *backpropagation* werd begin jaren tachtig nog afgewezen voor een leidende AI-conferentie en volgens Yann LeCun gebruikten onderzoekers toen zelfs codewoorden om te maskeren dat zij met neurale netwerken werkten.[72] Het kostte tijd voordat het belang van deze nieuwe benadering werd gezien. Jeff Hawkins zei in 2004 dat AI op het gebied van beeldherkenning nog steeds onderdeed voor de vaardigheden van een muis.[73]

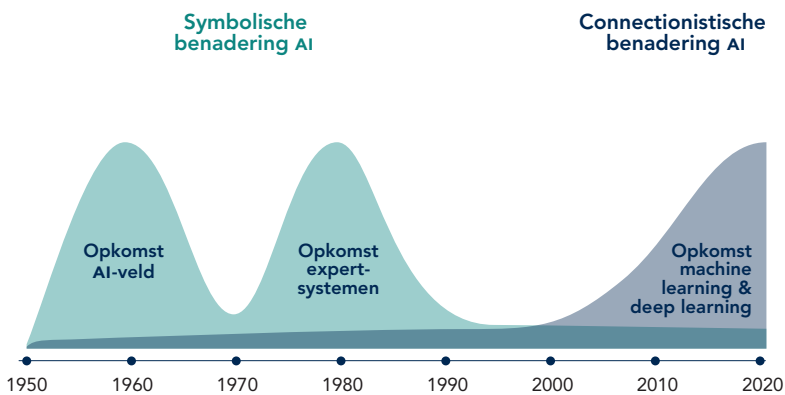
---

70 Russell 2019: 62-63.
71 Ihde 2010.
72 Uit een interview met Yann LeCun (Ford 2018: 122).
73 Tegmark 2017: 79.

Volgens schattingen uit die tijd zou het nog een eeuw duren voordat het mogelijk was om een mens te verslaan in het Chinese spel Go, dat veel meer combinaties aan zetten heeft dan schaken.[74]

Toch versloeg Google's AlphaGo al in 2016 de Go-wereldkampioen Lee Sedol. Dat kwam door recente doorbraken binnen de benadering van neurale netwerken. Onderzoekers als Yann LeCun en Andrew Ng speelden daarbij een belangrijke rol. Geoffrey Hinton wordt echter vaak gezien als de vader van die doorbraken. Samen met David Rumelhart en Ronald Williams had hij met een paper in *Nature* in 1986 al het gebruik van het *backpropagation algorithm* gepopulariseerd. Dat algoritme traceert de bijdrage van de outputlaag terug naar verborgen lagen daarachter. In die lagen worden de individuele units geïdentificeerd die aangepast moeten worden om het algoritme beter te laten functioneren. Lange tijd kende 'backprop' slechts een enkele verborgen laag. Sinds kort kunnen meerdere lagen onderscheiden worden. Daarmee adresseert *backpropagation* een centraal probleem van ANNs: het representeren van hiërarchie. Daardoor zijn relaties op verschillende niveaus te onderscheiden en wordt op alle niveaus bepaald wat het algoritme succesvol maakt (het toekennen van 'credit').[75] Dit soort neurale netwerken wordt sindsdien onder andere toegepast om de koers van aandelen op de beurs te simuleren. Figuur 1.7 toont de historische ontwikkeling van de twee benaderingen van AI.

Figuur 1.7 Verschuiving van een symbolische naar een connectionistische benadering van AI



74 Tonin 2019: 1.

75 In het boek *Perceptrons*, dat neurale netwerken diskwalificeerde als benadering, lieten Minsky en Papert zien dat de benadering geen oplossing kon bieden voor het vraagstuk van het zogenoemde XOR. Rumelhart, Hinton en Willians toonden dat backpropagation wel XOR kon leren.

In 1989 paste Yann LeCun backprop toe bij het trainen van neurale netwerken om handgeschreven postcodes te herkennen. Hierbij gebruikte hij *convolutional neural networks* (CNNs) waarbij complexe beelden worden opgeknipt in kleinere delen om de beeldherkenning efficiënter te maken. Ook dat was een belangrijke bijdrage aan hedendaagse AI-programma's.[76]

In een ander paper uit 2012 introduceerde Hinton het idee van 'dropout', dat het specifieke probleem van *overfitting* – het probleem van een model dat te zeer is toegespitst op de trainingsdata en daardoor niet meer goed werkt voor nieuwe data – in de training van neurale netwerken adresseerde. Hintons werk gaf een enorme impuls aan de toepasbaarheid van neurale netwerken in het veld van machine learning. Het gebruik van meerdere lagen in het trainingsproces is waarom hier gesproken wordt over 'deep' learning. Elke laag geeft op basis van de voorgaande een complexere representatie van de input. De eerste laag kan bijvoorbeeld hoeken en stippen identificeren. De tweede laag kan op basis daarvan onderdelen van een gezicht onderscheiden, zoals het puntje van een neus of de iris van een oog. De derde laag herkent neuzen en ogen, totdat een laag uiteindelijk het gezicht van een bepaald persoon herkent.[77]

Tekstbox 1.2 – Drie vormen van machine learning

Binnen machine learning worden drie verschillende vormen onderscheiden: *supervised, unsupervised* en *reinforcement learning*. Bij *supervised learning* wordt een programma data met labels gevoed, zoals in het eerdergenoemde voorbeeld 'kat' en 'niet-kat'. Het algoritme wordt op die data getraind en vervolgens wordt getest of het die labels op nieuwe data goed kan toepassen.

Bij *unsupervised learning* gebeurt dat niet en moet het algoritme zelf op zoek gaan naar patronen in de data. Het algoritme wordt hier met grote hoeveelheden ongelabelde data gevoed waaruit het zelf patronen gaat herkennen. Het uitgangspunt daarbij is dat eigenschappen die in de data geclusterd voorkomen, dat in de toekomst ook zullen doen. *Supervised learning* is bij uitstek geschikt wanneer duidelijk is waarnaar gezocht moet worden. Wanneer onderzoekers zelf nog niet goed weten wat voor patronen in data verborgen liggen en daar benieuwd naar zijn, is *unsupervised learning* geschikter.

76 Marcus en Davis 2019: 52.
77 Domingos 2015: 117.

De derde vorm is in weer andere contexten toepasbaar, zoals het spelen van een spel. Daar gaat het niet om een goed of fout antwoord of om het clusteren van data, maar draait het om strategieën die uiteindelijk kunnen leiden tot winst of verlies. Voor deze gevallen werkt de benadering van *reinforcement learning* beter. Het algoritme wordt hier getraind door het te belonen voor het volgen van bepaalde strategieën. De laatste jaren is *reinforcement learning* toegepast op allerlei klassieke computerspellen als Pacman en Atarispellen, evenals op 'gewone' kaartspellen en poker. Het algoritme krijgt als doel mee om de waarde van de score te optimaliseren en gaat vervolgens allerlei handelingen correleren met die score om een optimale strategie te ontwikkelen.

In 2012 won het team van Hinton een internationale wedstrijd op het gebied van *computer vision,* beeldverwerking met behulp van AI. Daarbij haalde het een foutmarge van 16 procent, terwijl daarvoor geen enkel team ooit minder dan 25 procent had behaald. Een aantal jaar eerder behaalde het team successen in spraakherkenning met neurale netwerken na een demonstratie van twee studenten in Toronto. Maar de winst met computer vision in 2012 was voor veel onderzoekers een openbaring.[78] Deep learning verspreidde zich en in 2017 hadden bijna alle teams in deze wedstrijd foutmarges lager dan 5 procent, vergelijkbaar met de score van mensen, en de verbetering van de foutmarge gaat door. De toepassing van deep learning is sindsdien in een sneltreinvaart geraakt. De wetenschappelijke doorbraken binnen de benadering van neurale netwerken zorgden voor een explosie aan activiteit in het AI-veld. Op dit moment bevinden we ons middenin deze AI-zomer. In het volgende hoofdstuk gaan we nader in op de ontwikkelingen die hierdoor in gang zijn gezet buiten het lab, in de markt en in de bredere samenleving.

Het is duidelijk dat de vlucht die AI de laatste jaren heeft genomen, zijn oorsprong vindt in fundamenteel wetenschappelijk onderzoek. Vervolgens haastten grote bedrijven als Google zich om getalenteerde onderzoekers binnen dit domein in dienst te nemen. Het waren echter wetenschappers aan universiteiten die verantwoordelijk zijn geweest voor de belangrijkste doorbraken in het veld.

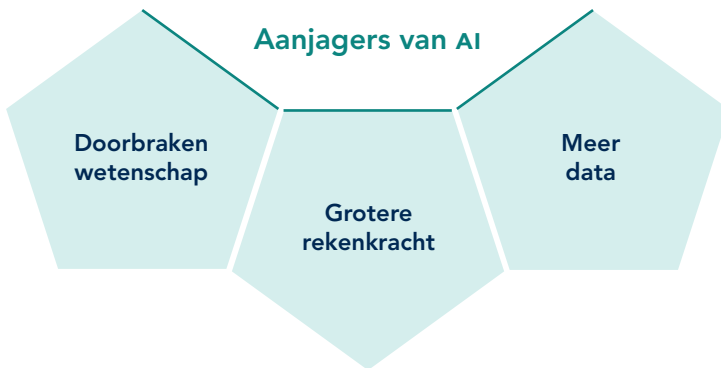
Naast deze academische mijlpalen liggen nog twee andere factoren ten grondslag aan de recente opkomst en toepassing van AI. De eerste daarvan is de groei in rekenkracht (*processing power).* Die groei wordt gevat met de Wet van Moore. Dit patroon wordt al decennia in de computerindustrie geobserveerd en

78 Uit een interview met Geoffrey Hinton (Ford 2018: 77).

beschrijft dat het aantal transistoren op een chip grofweg elke twee jaar verdubbelt. Steeds grotere rekenkracht wordt zo tegen lagere prijzen beschikbaar. Daardoor overtreffen de individuele smartphones van nu de rekenkracht van de beste computers van een paar decennia geleden. We merkten eerder op dat de eerste AI-winter mede te wijten was aan de combinatorische explosie. Grotere rekenkracht was daar een antwoord op. Een andere sprong in rekenkracht kwam uit de chipindustrie door het gebruik van *graphic processing units* (GPUs) in plaats van de klassieke *central processing units* (CPUs). GPUs waren oorspronkelijk ontwikkeld voor complexe afbeeldingen in de gaming-industrie, maar bleken vervolgens geschikt om parallel veel meer berekeningen met AI mogelijk te maken.[79] Vanaf 2015 wordt ook gebruik gemaakt van *tensor processing units* (TPUs) die speciaal ontworpen zijn voor machine learning-applicaties.

De andere factor die heeft bijgedragen aan de actuele AI-golf, is de toename van de hoeveelheid data. Die toename hangt sterk samen met de opkomst van het internet. De databronnen waarop algoritmes in het verleden toegepast konden worden, waren beperkt. Doordat mensen in de loop der jaren echter op steeds grotere schaal gebruik zijn gaan maken van het internet en met hun onlineactiviteiten direct en indirect veel digitale informatie genereren, is de hoeveelheid digitale data om te analyseren de afgelopen decennia sterk toegenomen.

Figuur 1.8 Drie aanjagers van de vooruitgang in AI



De 'digitale broodkruimels' die we achterlaten op het internet zijn dus voeding voor het trainen van AI-algoritmes. Maar we helpen het trainen nog verder. Door persoonsnamen te 'taggen' in foto's op Facebook voorzien mensen

79 Kelly 2017: 38.

algoritmes bijvoorbeeld van labels waarmee gezichtsherkenning kan worden getraind. Een specifieke dataset die van groot belang is voor de training van algoritmes is *Imagenet*, een open database van meer dan veertien miljoen met de hand gelabelde afbeeldingen. Ook het Internet of Things, de groei van het aantal sensoren en verbindingen in de fysieke omgeving, draagt bij aan de groei van data.

De driehoek van wetenschappelijke doorbraken, grotere rekenkracht en meer data heeft ervoor gezorgd dat AI recent een enorme vlucht kon nemen (figuur 1.8). Die vlucht wordt dus veelal gedreven door de toepassing van machine learning binnen de benadering van neurale netwerken – en binnen machine learning door de ontwikkeling van deep learning.

Kernpunten – AI in het lab

- In het lab heeft AI drie golven doorgemaakt. Daartussenin vonden twee winters plaats doordat de wetenschappelijke vooruitgang uitdoofde, hardwarecapaciteiten niet toereikend waren en hoge verwachtingen niet ingelost werden.
- De eerste golf van AI begon met de *Dartmouth Summer Research Project on AI* in 1956. Er volgden twee stromingen van symbolische AI en neurale netwerken. AI werd toen vooral toegepast op spellen als dammen, in vroege robots en voor wiskundige vraagstukken.
- De tweede golf begon in de jaren tachtig, mede gedreven door de internationale competitie tussen Japan, de VS en Europa. Deze golf bracht expertsystemen voort, de eerste grotere commerciële toepassing van AI.
- De derde golf begon in jaren negentig met wapenfeiten van symbolische AI, maar kwam later pas echt in een stroomversnelling met de vooruitgang binnen het domein van machine learning, en daarbinnen deep learning. Wetenschappelijke doorbraken op dit gebied vormden samen met de toename van rekenkracht en data de stuwende kracht achter deze huidige golf.

2. AI verlaat het lab en gaat de samenleving in

Sinds de geboorte van AI in 1956 hebben door de jaren heen verschillende toepassingen ervan het lab verlaten en zich in de samenleving verspreid. Expertsystemen worden al decennia op allerlei plekken ingezet en ook de eerste neurale netwerken vonden destijds hun weg naar de financiële sector. Grootschalige impact bleef tot op heden echter uit door de beperkte mogelijkheid om deze vormen van AI in te zetten.

Dat is nu anders. Door de recente stroomversnelling waarin AI is terechtgekomen, verspreiden zich nu voor het eerst allerlei toepassingen door de hele samenleving en economie. In het vorige hoofdstuk zagen we dat deze versnelling gedreven wordt door een combinatie van wetenschappelijke doorbraken, steeds grotere rekenkracht en een groeiende beschikbaarheid van data. In dit hoofdstuk bezien we hoe AI haar intrede doet in de samenleving. We brengen daartoe eerst een set van indicatoren in kaart die het huidige momentum van AI aantonen. Deze indicatoren variëren van publicaties en patenten tot investeringen en werkgelegenheid. Daarna bespreken we welke verschillende soorten AI er momenteel zijn, zoals beeldherkenning, spraakherkenning en robotica. Daarmee wordt duidelijk hoe breed AI-toepassingen inmiddels zijn verspreid, ook in Nederland. Vervolgens schetsen we hoe, vooral als gevolg van de intrede in de samenleving, het debat over AI op gang is gekomen. Ten slotte kijken we naar de toekomst van het lab. AI mag dan inmiddels de samenleving in zijn gegaan, juist bij deze technologie zal het lab van belang blijven.

2.1 Momentum van lab naar samenleving

Wetenschappelijke activiteit

Rond 2010 lijkt zich een momentum af te tekenen waarbij AI definitief en breed de overstap maakt van de wetenschappelijke wereld van het lab naar de praktijk van onze samenleving. Daaraan gaat een stijging van wetenschappelijke activiteit vooraf. De World Intellectual Property Organization publiceerde een studie waaruit blijkt dat het aantal AI-gerelateerde publicaties de laatste twintig jaar sterk is gestegen: tussen 1996 en 2001 jaarlijks met gemiddeld 8 procent, en tussen 2002 en 2007 met 18 procent.[80] Na 2015 nam die jaarlijkse groei opnieuw een spurt naar 23 procent, en in 2018 bedroeg het aandeel van AI-gerelateerde publicaties zo'n 2 tot 3 procent van het totale aantal gepubliceerde artikelen wereldwijd.[81] Nagenoeg een verdriedubbeling ten opzichte van de late jaren negentig.

80 World Intellectual Property Organization 2019.
81 Baruffaldi et al. 2020; Perrault et al. 2019.

Praktisch potentieel

De in het vorige hoofdstuk beschreven doorbraken in spraakherkenning en beeldherkenning door de techniek van deep learning openden in deze periode de deur naar allerlei praktische toepassingsmogelijkheden. Tegelijkertijd zien we dan ook dat het aantal toegekende patenten stijgt. Tussen 2006 en 2011 groeide het aantal AI-patenten jaarlijks met gemiddeld 8 procent en tussen 2012 en 2017 steeg die groei naar 28 procent.[82] In twee jaar tijd nam het aandeel dat AI-gerelateerde patenten uitmaakt van het totale aantal uitvindingen, toe van minder dan 1,5 procent naar bijna 2,5 procent in 2017.[83] De helft van alle AI-uitvindingen die ooit geregistreerd zijn, werden gepatenteerd tussen 2013 en 2018.[84] Kortom, vanaf de vroege jaren tien van deze eeuw gaat de wetenschappelijke stroomversnelling hand in hand met een golf aan AI-patenten.

Kijken we naar de gehonoreerde patenten, dan is de toename het grootst op het terrein van machine learning. In 40 procent van de totale toegekende patenten wordt hieraan gerefereerd. Binnen dat gebied is de stijging weer het hoogst voor deep learning: het aantal patenten op dit gebied steeg tussen 2013 en 2016 met 175 procent.[85] Als we kijken naar de toepassingsgebieden die we in de volgende paragraaf uitgebreider bespreken, dan is beeldverwerking of computer vision de grootste categorie, met ongeveer de helft van alle patenten in die periode.[86] Er wordt dus volop geïnnoveerd met AI. Het toenemende belang van daadwerkelijke toepassingen blijkt ook uit de ontwikkeling van AI-software: vanaf 2014 groeide de ontwikkeling van open source software (OSS) op het gebied van AI drie keer zo hard vergeleken met andere OSS.[87]

Stijgende investeringen: AI wordt business

De toename van het aantal patenten reflecteert een groeiende interesse van het bedrijfsleven in AI. Vanaf ongeveer 2010 werken bedrijven als Google, IBM en Microsoft met neurale netwerken voor spraakherkenning. Google past deze neurale netwerken sinds 2012 toe op Android smartphones. Ook het gebruik van computer vision bij grote technologiebedrijven groeit sindsdien. In 2014 nam Google het Britse DeepMind over, dat wereldwijd leidend onderzoek op het gebied van AI doet en allerlei complexe doelstellingen als eerste wist te bereiken, met de Go-overwinning op Lee Sedol als bekendste mijlpaal.

---

82 World Intellectual Property Organization 2019.
83 Baruffaldi et al. 2020.
84 World Intellectual Property Organization 2019: 39.
85 World Intellectual Property Organization 2019; Baruffaldi et al. 2020.
86 World Intellectual Property Organization 2019; Baruffaldi et al. 2020.
87 Baruffaldi et al. 2020:32.

Verbeteringen van AI op het gebied van vertalingen werden vanaf 2016 toegepast voor Google Translate.[88] In 2017 kocht Intel voor 14 miljard euro het Israëlische bedrijf Mobileye dat gespecialiseerd is in bestuurders ondersteunende en autonome rijsystemen. Ook Facebook, Amazon, Apple, Microsoft en andere hard- en softwarebedrijven namen de afgelopen jaren AI-start-ups over om hun AI-capaciteiten te versterken. Waar er in 2010 nog krap tien acquisities werden geregistreerd, waren dat er in 2019 ruim 240.[89]

Grote technologiebedrijven namen ook prominente AI-wetenschappers in dienst. Geoffrey Hinton ging werken bij Google, Yann LeCun bij Facebook en Andrew Ng werkte bij Google en het Chinese bedrijf Baidu. In publieke uitingen benadrukken de leiders van deze techbedrijven expliciet hun aandacht voor AI. Jeff Bezos van Amazon schreef in 2016 in een brief aan aandeelhouders dat machine learning cruciaal was om kernoperaties te verbeteren. De CEO van Google Sundar Pichai zei in een speech in 2017 dat het bedrijf verschoof van een "mobile-first world" naar een "AI-first world".[90] In een e-mail aan de werknemers in 2018 kondigde Satya Nadella van Microsoft organisatorische veranderingen aan om de middelen meer op de cloud (online opslag) en AI te richten.[91] Ook Chinese techgiganten als Baidu, Tencent en Alibaba geven al jaren – en in sommige gevallen zelfs eerder dan hun Amerikaanse tegenhangers[92] – aan dat AI de kern van hun businessstrategieën uitmaakt. Zo was Alibaba's eerste onderzoekscentrum buiten China er een in Singapore gericht op AI.

Naast deze 'Big Tech' zijn er ook allerlei gespecialiseerde jongere bedrijven die AI in het hart van hun operatie hebben, zoals Waze, Shazam, Face++, Tesla, Spotify en Booking.com. De ambitie van de Europese Commissie is dat in 2030 drie op de vier bedrijven gebruikmaken van AI.[93] De meeste grote Nederlandse bedrijven zijn dat volgens een studie ook van plan of zijn daar al mee bezig.[94] Nederland kent ook innovatieve start-ups die AI gebruiken, zoals Pacmed en Seedlink, die zich respectievelijk richten op het gebruik van AI in de zorg en voor HRM.

Wereldwijd groeien de investeringen in AI-start-ups de laatste jaren gestaag. Een onderzoek van Stanford University schat de totale private investering in AI-start-ups in 2018 op 40 miljard USD, waar dat in 2010 nog 1,3 miljard was.

---

88 Uit een interview met Yoshua Bengio (Ford 2018: 27-28).
89 CB Insights, 24 juni 2021.
90 Agrawel et al. 2018: 179.
91 Leung 2019: 248.
92 CB Insights, 26 april 2018.
93 Europese Commissie, 9 maart 2021.
94 Denkwerk 2018.

De investeringen namen in die tijd jaarlijks met gemiddeld bijna 50 procent toe.[95] Hoewel de omvang van de schattingen afhankelijk is van definities en methodologische keuzes, is de opwaartse trend onmiskenbaar. Niet alleen de totale omvang maar ook het aantal investeringen nam toe: van 200 in 2011 tot 1.400 in 2017. Uit die ontwikkelingen leidt de OESO af dat de potentie van AI doordringt bij investeerders.[96]

Kijken we breder dan start-ups, dan schat Stanford University de totale investering in AI-bedrijven op bijna 70 miljard USD in 2020.[97] Dat is vijf keer zoveel als in 2015. Tussen 2015 en 2020 kregen AI-bedrijven wereldwijd dus een aanzienlijke financiële injectie. In de laatste jaren is 60 procent van alle investeringen in AI naar machine learning gegaan.[98] Lang was het grootste deel van die investeringen gericht op de ontwikkeling van autonome voertuigen, in lijn met de eerdergenoemde focus op computer vision.[99] In 2018 was 30 procent van de investeringen in AI-start-ups gericht op de ontwikkelingen van autonome voertuigen en in Californië verzevenvoudigde het aantal bedrijven dat dit soort voertuigen test. De COVID-19-pandemie lijkt in 2020 echter een verschuiving teweeg te hebben gebracht, waardoor de meeste investeringen in dat jaar gericht waren op de gezondheidssector en medicijnontwikkeling.[100]

Impact op economie en werk

Verschillende consultancybureaus wagen zich aan voorspellingen over de implicaties van de definitieve intrede van AI in de samenleving. Ze voorzien dat AI door haar generieke karakter invloed op vrijwel alle bedrijfssectoren zal hebben, en een aanzienlijke impact op de economie. PwC stelt in 2017 dat AI in 2030 tot wel 15,7 biljoen USD zou kunnen bijdragen aan de wereldeconomie.[101] In datzelfde rapport merkt PwC de zorg, de automobielindustrie, financiële dienstverlening, transport en logistiek, ICT en media, en retail aan als sectoren waarbinnen de komst van AI veel effect zal sorteren. Ook Deloitte voorziet een snel toenemend belang van AI voor het bedrijfsleven en stelt dat het tijdsbestek waarbinnen bedrijven hun kans kunnen grijpen om met AI competitief voordeel te behalen, zeer kort is. Bedrijven moeten dus snel instappen om de boot niet te missen.[102] In een rapport uit 2018 voorspelt McKinsey dat 70 procent van de bedrijven wereldwijd gebruik zal maken van AI en dat de technologie de

95 Baruffaldi et al. 2019: 82.
96 Baruffaldi 2020: 1.
97 Zhang et al. 2021: 93.
98 Tonin 2019.
99 Baruffaldi et al. 2019: 90; OESO 2018: 3.
100 Zhang et al. 2021: 97.
101 Rao en Verweij 2017.
102 Loucks et al. 2019.

potentie heeft om het mondiale bruto binnenlands product (bbp) met 1,2 procent per jaar te doen groeien.[103] In een recenter rapport analyseert McKinsey de economische potentie van AI voor zogenoemde Europese 'digitale koplopers', waaronder Nederland. Als het deze landen lukt om AI succesvol te adopteren, kan de economische groei oplopen tot 1,4 procent van het bbp per jaar mits hierin overtuigend wordt geïnvesteerd, aldus de analisten.[104]

Amerikaans onderzoek laat verder zien dat de intrede van AI in de samenleving ook effect heeft op de werkgelegenheid. Het aantal AI-banen groeide van 0,3 procent in 2012 naar 0,8 procent van het totaal aantal vacatures in de VS in 2019. Het aandeel AI-gerelateerde banen steeg van 0,26 procent in 2010 naar 1,32 procent in 2019.[105] AI is een van de populairste onderwerpen geworden voor promovendi in de computerwetenschappen in Noord-Amerika. In 2010 was het aandeel AI-promovendi dat een baan kreeg in de industrie nog redelijk gelijk aan dat binnen de academie. Sindsdien is er een stijging te zien van het aandeel AI-promovendi dat een baan krijgt in het bedrijfsleven: in 2019 ging meer dan de helft van de AI-promovendi het bedrijfsleven in, ten opzichte van een krappe 24 procent die een academische carrière voortzetten.[106] Volgens technologie-expert Tim O'Reilly is 'data scientist' inmiddels de meest begeerde baantitel in Silicon Valley. Het McKinsey Global Institute schat dat de VS in 2018 al 140.000 tot 190.000 meer machine learning experts nodig had dan er beschikbaar waren.[107]

Ook overheden zetten in op AI

AI komt overigens niet alleen via bedrijven en toepassingen in de private sector de samenleving in, ook allerlei publieke diensten dragen daaraan bij. Politiediensten gebruiken de technologie bij opsporing en de bestrijding van criminaliteit, instanties op het gebied van de sociale zekerheid gebruiken haar om fraude te detecteren en tijdens de coronacrisis ontstonden er initiatieven om met behulp van AI de pandemie aan te pakken. Een wereldwijd en historisch overzicht ontbreekt, maar de Europese Commissie telde in 2019 grofweg 230 toepassingen van AI in de publieke sector.[108] Het is heel waarschijnlijk dat het daadwerkelijke aantal hoger ligt; in datzelfde jaar telde TNO in Nederland al 74 projecten met AI in de publieke dienstverlening.[109]

---

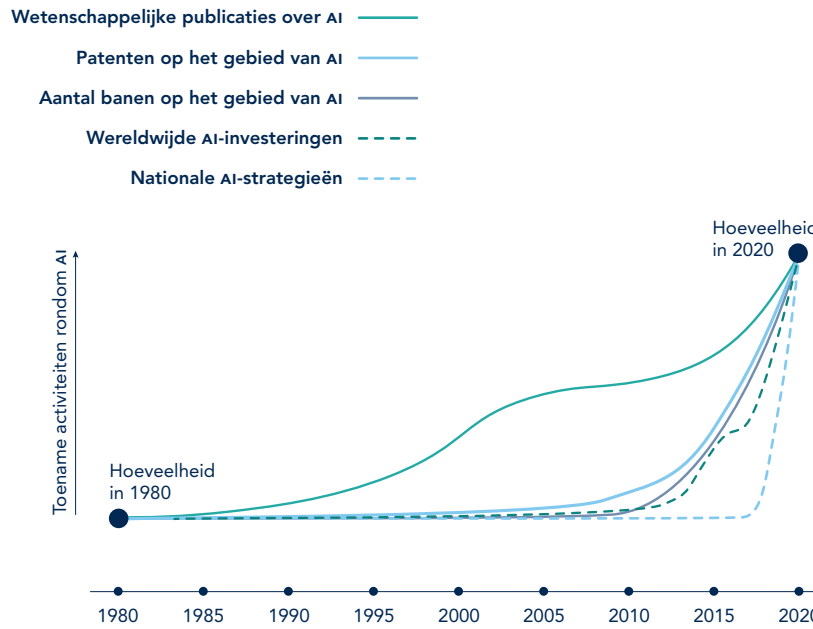
103 Bughin et al. 2018
104 McKinsey & Company 2020.
105 Perrault et al. 2019.
106 Zhang et al. 2021: 118.
107 Domingos 2015: 9.
108 Misuraca en Van Noordt 2020.
109 Van Veenstra et al. 2019.

Dat AI de laatste jaren momentum krijgt in de samenleving, blijkt ook als we naar een andere indicator kijken: nationale AI-strategieën. Nadat duidelijk werd dat AI op een punt was aangekomen dat er allerlei praktische toepassingen in het verschiet lagen en het bedrijfsleven flink is gaan investeren, hebben ook overheden strategieën ontwikkeld om de vruchten van de technologie te kunnen plukken. In maart 2017 was Canada de eerste, met de *Pan-Canadian Artificial Intelligence Strategy*. Daarin kondigde de Canadese regering een investering in AI aan van 125 miljoen dollar. Datzelfde jaar volgden ook Singapore, Japan en de Verenigde Arabische Emiraten met AI-strategieën. China publiceerde het *New Generation Artificial Intelligence Development Plan*, waarin het de ambitie uiteenzet om in 2030 de absolute wereldleider te zijn. Snel volgen landen als Finland, de VS, Frankrijk, het Verenigd Koninkrijk en Duitsland. Ook de EC begint in het kader van 'een Europa dat klaar is voor het digitale tijdperk' meerdere trajecten rondom AI, met een Europees actieplan voor AI[110] en een datastrategie[111]. Sindsdien hebben tientallen landen een actieplan opgesteld om aan de slag te gaan met AI, waaronder ook verrassende landen als Kenia, India en Mexico.[112] Na een piek van twintig publicaties in 2019 staat de teller inmiddels op zo'n 60 nationale AI-strategieën. Eind 2019 presenteerde staatssecretaris Mona Keijzer de Nederlandse strategie: het *Strategisch Actieplan voor AI* (SAPAI).[113]

In het kielzog van de wetenschappelijke stroomversnelling waarin AI vanaf 2000 terechtkomt, indiceren de toename van het aantal patenten en de private investeringen, de opkomst van nieuwe businessmodellen, de groei van AI-gerelateerde banen en de lancering van nationale AI-strategieën een nieuwe fase in de geschiedenis van AI: de technologie gaat de samenleving in. Figuur 2.1 geeft dit momentum weer aan de hand van de besproken indicatoren. De volgende vraag is: hoe gebeurt dat? Welke gedaanten neemt AI aan in de samenleving? Daarover gaat de volgende paragraaf.

110 Europese Commissie 2021a [2018].
111 Europese Commissie 2020
112 HolonIQ, 9 april 2020; Future of Life Institute, z.d (a); Van Roy et al. 2021.
113 Ministerie van Economische Zaken en Klimaat, Ministerie van Justitie en Veiligheid, Ministerie van Sociale Zaken en Werkgelegenheid, Ministerie van Onderwijs, Cultuur en Wetenschap, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties 2019.

Figuur 2.1 AI krijgt momentum buiten het lab



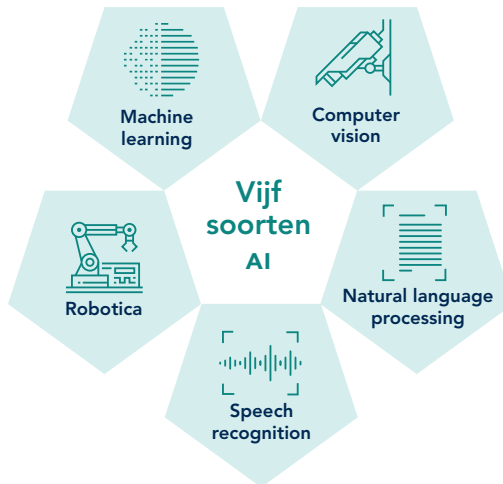
Kernpunten – Momentum van lab naar samenleving

- Vanaf de jaren tien van deze eeuw krijgt AI momentum in de samenleving. De eerdere doorbraken in het lab vormen een springplank om AI in de praktijk toe te passen.
- Het nieuwe praktische potentieel van AI weerspiegelt zich onder andere in een toename van het aantal patenten. De helft van alle AI-uitvindingen is gepatenteerd tussen 2013 en 2018.
- Grote techbedrijven gaan zich openlijk toeleggen op AI, er ontstaan nieuwe bedrijven met AI in het hart van de operatie en wereldwijd stijgen de private investeringen aanzienlijk.
- Door het generieke karakter wordt een grote impact op de economie voorspeld. Op de arbeidsmarkt groeit de vraag naar AI-experts en meer promovendi kiezen voor een baan in het bedrijfsleven.
- Ook overheden gaan inzetten op AI en inmiddels hebben meer dan 60 landen een nationale AI-strategie ontwikkeld.

2.2 AI in de praktijk

AI heeft dus de stap van het lab naar de samenleving gemaakt. Dat betekent dat we de technologie in allerlei toepassingen tegenkomen: in de vorm van chatbots, slimme camera's, vertaalapps, aanbevelingssystemen, risicoanalyses, rijsystemen en ga zo maar door. AI is in de praktijk veelvormig. We kunnen een aantal soorten onderscheiden, grofweg op basis van het type taak dat wordt uitgevoerd. In het veld worden daarvoor verschillende indelingen gehanteerd. Voor het overzicht maken we hier onderscheid tussen vijf soorten AI: toepassingen die gericht zijn op het doen van voorspellende analyses (*machine learning*), op het verwerken van beeld (*computer vision*), taal (*natural language processing*) en spraak (*speech recognition*), en het uitvoeren van fysieke handelingen (robotica). Deze toepassingen zien we op dit moment al om ons heen. Figuur 2.2 geeft een overzicht van de vijf soorten AI die we hierna bespreken.

Figuur 2.2 Vijf soorten AI in de praktijk



Machine learning

De meest voorkomende soort AI is machine learning. Dat is enigszins verwarrend omdat dezelfde term wordt gebruikt voor de op dit moment dominante techniek binnen AI. De term 'machine learning' wordt echter ook gebruikt om te verwijzen naar een bepaald type toepassing gericht op het doen van voorspellende analyses (*predictive analytics* of *advanced analytics*). Met behulp van machine learning wordt dan gezocht naar patronen in datasets, om daaruit vervolgens voorspellingen af te leiden. Hoewel de techniek van machine learning ook bij andere soorten AI kan worden gebruikt, is de voorspelling zelf in dit geval de hoofdtaak. We zouden dit soort AI daarom ook wel 'voorspelsystemen' kunnen noemen.

De mogelijkheid om op basis van data beter geïnformeerde inschattingen te maken over de toekomst kan een grote meerwaarde hebben bij tal van activiteiten. Neem het organiseren van de energietoevoer. Google's DeepMind ontwikkelde een AI-systeem dat op basis van weersvoorspellingen en turbinedata 36 uur vooruit kan voorspellen hoeveel windenergie er binnenkomt.[114] Ondanks de variabiliteit van de wind is het zo mogelijk optimaal gebruik te maken van windenergie.

Omdat risicovoorspellingen van oudsher een belangrijke rol spelen in de financiële dienstverlening komen veel voorbeelden van machine learning uit deze sector. AI ondersteunt hier bijvoorbeeld kredietbeoordeling, risicomanagement en fraudedetectie. Kredietbeoordelaars als Experian en Focum beheren grote databases met gegevens van zowel bedrijven als miljoenen Nederlanders. Door verschillende gegevens met elkaar te combineren, zoals de kredietthistorie en persoonlijke gegevens, komen de beoordelaars met behulp van machine learning tot een voorspelling over iemands kredietwaardigheid. Maar er zijn ook AI-toepassingen voor de klanten zelf. Zo biedt ING AI-applicaties aan die financiële trends voorspellen.[115] Klanten krijgen daarmee inzicht in ontwikkelingen die impact kunnen hebben op hun werkkapitaal en kunnen zich zo informeren bij beleggingskeuzes.

Voorspellingen op basis van machine learning kunnen ook waardevol zijn bij het tegengaan van fraude of criminaliteit. Onder andere banken, verzekeraars, gemeenten en de politie verkennen de mogelijkheden om zich via machine learning te informeren bij de opsporing en bestrijding van misstanden. Banken brengen met AI bijvoorbeeld het betaalpatroon van een klant in kaart. Een transactie die afwijkt van dat patroon, kan reden zijn voor de bank om een extra controle uit te voeren. Ook gemeenten gebruiken machine learning om zicht te krijgen op mogelijke fraudegevallen. Een veelbesproken voorbeeld dat uitgebreider aan bod komt in hoofdstuk 6, is het Systeem Risico Indicatie (SyRI) dat als doel heeft om bijstandsfraude tegen te gaan. Maar er zijn ook gemeenten, zoals Amsterdam en Twenterand, die AI gebruiken om te voorspellen welke inwoners in grote schulden dreigen te raken en dus hulp kunnen gebruiken om uit de problemen te blijven.

AI in de vorm van voorspelsystemen is ook terug te vinden bij de politie. Veel voorbeelden van *'predictive policing'* komen uit de VS, waar AI onder andere is ingezet om de kans op recidive bij gedetineerden te voorspellen. Ook in Nederland gebruikt de politie machine learning om zich te informeren. Zo

114 DeepMind, 26 februari 2019.
115 ING's Smart Working Capital Assistant en Katana Lens.

brengt ze met het Criminaliteit Anticipatie Systeem (CAS) patroonmatige criminaliteit in kaart en voorspelt ze waar en wanneer de kans op een overval het grootst is. Agenten gebruiken deze informatie om daarop te anticiperen door bijvoorbeeld op een bepaalde plek extra te gaan surveilleren. Op vrijwel alle 168 politiebureaus in Nederland wordt een versie van CAS gebruikt.[116] De politie experimenteert daarnaast met machine learning om te voorspellen welke cold cases de hoogste kans hebben op een doorbraak en dus zinvol zijn om verder te onderzoeken.[117]

Niet alleen de financiële dienstverlening, fraudebestrijding en politie zijn gebaat bij een grotere accuraatheid van voorspellingen met behulp van AI. Dat geldt ook voor supermarkten, die hebben aangekondigd te gaan experimenteren met dynamische prijzen om voedselverspilling tegen te gaan en de omzet te optimaliseren. Albert Heijn zet machine learning bijvoorbeeld in om geautomatiseerd kortingen te bepalen. Het algoritme maakt daarvoor gebruik van gegevens over de houdbaarheid van het product, maar neemt ook de winkellocatie, de weersomstandigheden en het historische verkoopverloop mee in de berekening.[118]

In de media-industrie wordt machine learning gebruikt om producten en diensten beter af te stemmen op consumenten. De bekendste voorbeelden zijn platformdiensten als Netflix, Youtube en Spotify, die AI gebruiken om op basis van kijk- en luisterpatronen zo relevant mogelijke aanbevelingen te doen. Dit soort voorspelsystemen wordt daarom ook wel *'recommender systems'* genoemd. Personalisering met behulp van machine learning heeft zich eveneens ontwikkeld tot een belangrijke pijler in de e-commerce. Het maakt de kern uit van onlinewinkels als Amazon, Bol.com en Booking.com, die met behulp van AI gebruikersprofielen opstellen en daarop hun marketing afstemmen.

Op een vergelijkbare manier kunnen ook advertenties zo nauwkeurig mogelijk worden afgestemd op de interesses en gevoeligheden van individuele gebruikers (*microtargeting*). Microtargeting kan commercieel worden ingezet, maar ook voor politieke doeleinden. Politieke microtargeting kwam wereldwijd in opspraak toen bekend werd dat Cambridge Analytica ten tijde van de Amerikaanse presidentsverkiezingen en het Brexit-referendum in 2016 gericht politieke advertenties verspreidde op basis van Facebookdata (zie tekstbox 2.1). Microtargeting is echter een breed fenomeen dat ook in Nederland al jaren voorkomt.[119] Uit onderzoek van Argos en de Groene Amsterdammer blijkt

116 Waardenburg et al. 2020: 70.
117 Politie, 23 mei 2018.
118 Albert Heijn, 20 mei 2019.
119 Prins 2017.

dat vrijwel alle politieke partijen online op maat gemaakte boodschappen verspreiden.[120]

Tekstbox 2.1 – Cambridge Analytica en microtargeting
Microtargeting is het toespitsen van *bepaalde* berichten op *bepaalde* mensen. Met behulp van AI kan 'psychografie' worden bedreven, waarbij de content die aan iemand wordt aangeboden zoveel mogelijk wordt aangepast aan diens profiel. Op die manier kan content gericht worden verspreid, om zo het effect ervan te optimaliseren. De meest bekende voorbeelden van de donkere kant van targeting zijn verbonden aan het datamining bedrijf Cambridge Analytica. Dit bedrijf speelde een belangrijke rol bij zowel de pro-Trump- als de pro-Brexit-campagne. Met behulp van machine learning en een enorme hoeveelheid gegevens over het online gedrag van mensen, leerde het bedrijf het publiek zo goed kennen dat het met gerichte boodschappen via onder andere Facebook het stemgedrag van mensen trachtte te beïnvloeden. Cambridge Analytica is inmiddels failliet, maar op dit moment wordt er een nieuw soort voorspelsystemen ontworpen – zogenoemde *multi-agent artificial intelligence* (MAAI). Van dit soort voorspelsystemen wordt geclaimd dat ze nog accurater het gedrag van mensen kunnen voorspellen én beïnvloeden, door targeting te testen in volledig gesimuleerde samenlevingen.[121]

Computer vision

Een ander type AI is toegespitst op beeldherkenning. Bij computer vision, zoals dit soort AI heet, wordt het vermogen tot *waarnemen, analyseren en interpreteren van visuele informatie* geautomatiseerd. Daarbij kan het gaan om foto's of video's, maar ook om de fysieke omgeving. De ontwikkeling van computer vision heeft een impuls gekregen door de toename van digitaal beeldmateriaal. Sociale media en smartphones hebben bijgedragen aan een ware explosie van (publiekelijk) beschikbare beelden die kunnen worden gebruikt om computer vision-algoritmes te trainen. Bovendien bestaat een toenemend deel van onze communicatie uit beeldmateriaal – *'pics, or it didn't happen'*. Sinds de oprichting van Instagram zijn er 50 miljard foto's geüpload op het platform, dagelijks

120 Davidson en Delhaas, 22 april 2020. De Wet op Politieke Partijen moet regels hiervoor opstellen (Rijksoverheid, 26 juni 2019).
121 Hern, 30 juli 2019; Lewis en Hilder, 23 maart 2018; Lawton, 2 oktober 2019.

worden er 350 miljoen foto's gepost op Facebook en elke minuut komt er 500 uur aan videomateriaal bij op YouTube.[122]

Een van de bekendste toepassingen van computer vision is gezichtsherkenning. Dat gaat een stap verder dan gezichtsdetectie: in het laatste geval is de computer in staat om een gezicht te detecteren, waar het bij gezichtsherkenning in staat is om een specifiek gezicht te herkennen. Via een camera worden allerlei gegevens op millimeterschaal afgelezen van het gezicht, zoals de vorm van de kin, afstand tussen de ogen en rondheid van de wangen. De computer vertaalt die data naar een 'gezichtscode' die unieke kenmerken representeert en op basis daarvan het gezicht de volgende keer herkent.

Gezichtsherkenningssoftware is ingebouwd in sommige smartphones, waardoor gebruikers hun telefoon kunnen ontgrendelen door alleen naar de camera te kijken. Het werkt hier dus bij wijze van wachtwoord. Op eenzelfde manier werken verschillende apps met gezichtsherkenning, waardoor het voor betalingen met de telefoon bijvoorbeeld niet meer nodig is een pincode in te toetsen.

Volgens een inventarisatie van de Autoriteit Persoonsgegevens wordt gezichtsherkenning het meest ingezet bij bedrijven in de detailhandel, beveiliging, sport, entertainment, vervoer en bij gemeenten.[123] Enkele gemeenten gebruiken gezichtsherkenning om identiteitsfraude tegen te gaan bij de aanvraag van nieuwe identiteitsdocumenten.[124] In de *Agenda Digitale Overheid* kondigde het kabinet aan om vanaf 2020 het gebruik van gezichtsherkenning in het publieke domein verder te verkennen en daarmee te experimenteren.[125] Tegelijkertijd verbindt de EU strenge regels aan het gebruik van gezichtsherkenning in de openbare ruimte: in de recent voorgestelde AI-Verordening wordt dit in principe verboden tenzij er zwaarwegende redenen zijn om dit middel in te zetten.[126]

China loopt wereldwijd voorop als het gaat om overheidsgebruik van gezichtsherkenning. De technologie wordt daar onder andere ingezet door de politie en om de openbare ruimte in de gaten te houden.[127] Maar ook veel Amerikaanse overheidsorganisaties als de politie, inlichtingendiensten en grensbewaking

122 Apple, 24 juni 2020; Smith, 18 september 2013; Wojciki, 14 februari 2020.
123 Autoriteit Persoonsgegevens, 29 oktober 2020.
124 Waarlo en Verhagen, 27 maart 2020.
125 Overheidsbreed Beleidsoverleg Digitale Overheid, 2020 [2018].
126 Europese Commissie 2021b.
127 Chen, 12 oktober 2017; Simonite, 3 september 2019, 3. Eerder werd ook bekend dat de Chinese overheid gezichtsherkenning gebruikt om Oeigoeren, een islamitische minderheid in China, op te sporen en in de gaten te houden (Mozur, 14 april 2019).

gebruiken gezichtsherkenning.[128] Hoewel er in Europa nog druk wordt gediscussieerd over hoe met deze technologie moet worden omgegaan, experimenteren ook hier de meeste lidstaten al veel met gezichtsherkenning op onder andere vliegvelden, in stadions, scholen, casino's en bij de politie.[129] Volgens AlgorithmWatch, een onderzoeks- en belangenorganisatie die zich op het gebruik van algoritmes en AI richt, gebruiken politieorganisaties uit zeker elf Europese landen gezichtsherkenning.[130]

In Nederland wordt onder andere in de Johan Cruijff Arena in Amsterdam een pilot gedaan met gezichtsherkenning. Bezoekers worden via slimme camera's gekoppeld aan hun stoelnummer. Zodra er ongeregeldheden plaatsvinden, is het mogelijk de camerabeelden aan te vullen met persoonsgegevens zodat het mogelijk is de overtreders op te sporen. Ook op Schiphol wordt gebruik gemaakt van gezichtsherkenning. Bij 'e-gates' wordt het gezicht van een reiziger gescand en automatisch vergeleken met de foto op het identiteitsbewijs. Sinds 2019 loopt er ook een proef op de luchthaven waarbij passagiers helemaal geen identiteitsbewijs meer hoeven te laten zien bij het boarden en kunnen inchecken met hun gezicht.[131] De Koninklijke Marechaussee experimenteert op Schiphol ook met computer vision om mensen te monitoren en afwijkend gedrag te detecteren zodat daar tijdig op geanticipeerd kan worden.

Computer vision wordt voor veel meer toepassingen gebruikt dan alleen gezichtsherkenning. Dit soort AI is ook cruciaal bij autonome voertuigen. (Semi-)zelfrijdende auto's die ontwikkeld worden door Tesla, BMW, Volvo, Audi en Uber, zijn uitgerust met meerdere camera's om de omgeving in kaart te brengen en daarbinnen objecten, wegmarkeringen, verkeersborden en stoplichten te herkennen. Bij andere toepassingen van computer vision is het monitoren van de fysieke omgeving het hoofddoel. Het ministerie van Infrastructuur en Waterstaat gebruikt dit soort AI bijvoorbeeld om te signaleren wanneer er (preventief) onderhoud nodig is aan wegen of kunstwerken om hogere kosten of zelfs ongelukken te voorkomen. De gemeente Amsterdam gebruikt computer vision om verontreiniging van lucht en straten tegen te gaan. Camera's detecteren de kentekens van voertuigen die een milieuzone binnenrijden en sturen gegevens van overtreders naar het Centraal Justitieel Incassobureau. Daarnaast wordt er geëxperimenteerd met camera's op vuilniswagens die zijn getraind

128 De belangenorganisatie Fight for the Future, die zich inzet voor digitale rechten, houdt op een kaart bij waar de Amerikaanse overheid gezichtsherkenningstechnologie toepast (Fight for the Future z.d.).
129 Chiusi et al. 2020.
130 Kayser-Bril, 11 december 2019.
131 Passagiers dienen zich eenmalig te registreren bij een kiosk met hun paspoort en boarding pass, waarna ze de rest van de reis automatisch worden geïdentificeerd (Schiphol, 18 februari 2019).

om afval te herkennen, waarna gericht actie kan worden ondernomen tegen bijvoorbeeld verkeerd aangeboden vuilnis. Amsterdam zette computer vision ook in met de 'anderhalvemeter-monitor' om drukte te detecteren en mensen te stimuleren afstand te houden in verband met het coronavirus.[132]

Naast dit soort toepassingen in de publieke ruimte heeft computer vision ook bijzonder veel potentie voor de landbouwsector en voedselindustrie (zie tekstbox 2.2). De Nederlandse start-up OneThird heeft een groente- en fruitscanner ontwikkeld om de houdbaarheid op de dag nauwkeurig te kunnen inschatten met behulp van beeldherkenning.[133] Daardoor kunnen slimmere keuzes worden gemaakt om voedselverspilling tegen te gaan, bijvoorbeeld door ervoor te kiezen om een lading tomaten bij aankomst in het distributiecentrum direct te versnijden in een salade omdat ze bij aankomst in de winkel niet vers genoeg meer zouden zijn. In Nederland onderzoeken onder andere het Agro-Food lab in Wageningen en het ministerie van Landbouw, Natuur en Voedselkwaliteit (LNV) hoe AI van meerwaarde kan zijn in de land- en tuinbouw en de veehouderij. Sinds 2018 stimuleert LNV de adoptie van onder andere AI voor zogenoemde 'precisielandbouw', om daarmee een stap te zetten richting kringlooplandbouw.[134]

Hoewel er binnen het veld van computer vision vorderingen worden gemaakt, zijn de toepassingen vaak nog beperkt tot een specifiek domein en daarbinnen tot een specifieke taak. Dat is de reden dat computer vision relatief succesvol is in het interpreteren van medische beelden[135]: het gaat bij de beeldgebaseerde diagnostiek om het vergelijken van specifieke beelden op de onregelmatigheden die kunnen wijzen op een aandoening. Het algoritme kan radiologen, dermatologen en pathologen in dat geval ondersteunen om zulke afwijkingen te detecteren.[136] Bovendien is de gezondheidszorg een datarijke sector, waarvan een groot deel visueel; er zijn dus veel beelden om systemen mee te trainen.

---

132 Amsterdam Algoritmeregister, z.d.
133 OneThird, z.d.
134 Nationale Proeftuin Precisie Landbouw, z.d.
135 Uit een meta-analyse blijkt dat de diagnostische prestaties van deep learning-systemen vergelijkbaar zijn met die van medische professionals (Liu et al. 2019). De auteurs plaatsen daarbij echter de kanttekening dat in de onderzochte studies de externe validatie doorgaans ontbrak en de resultaten om die redenen niet de algemene conclusie rechtvaardigen dat AI even goed presteert in de op beeld gebaseerde diagnostiek als artsen.
136 Yu et al. 2018.

Tekstbox 2.2 – AI in de landbouw

De laatste decennia zijn er veel AI-toepassingen ontwikkeld die de agrarische bedrijfsvoering 'data-gedreven' ondersteunen. Een mooi voorbeeld is *Ida*: een door een Nederlands bedrijf ontwikkeld systeem voor melkveehouders. Ida maakt gebruik van sensoren in een halsband om de bewegingen van een koe vast te leggen. Het systeem herkent afwijkende bewegingspatronen als indicaties van, bijvoorbeeld, ziekten of tochtigheid en raadt de boer daarbij passende handelingen aan. Zo kunnen ziekteduur en antibioticagebruik worden teruggedrongen en kan de melkopbrengst worden geoptimaliseerd. Doordat Ida boeren vraagt haar inschattingen te verifiëren en logboeken bij te houden, worden data vergaard waarmee het herkenningssysteem door middel van *supervised learning* vervolgens weer wordt getraind.

De laatste jaren wordt verdere automatisering van het boerenbedrijf nagestreefd. Nederland is een pionier op dit gebied; 30 jaar geleden werden bijvoorbeeld de eerste Nederlandse melkrobots ontwikkeld. Zo'n robot heeft sensoren die de uiers van koeien herkennen waardoor de melkzuigers automatisch bevestigd kunnen worden. Tijdens het melken worden vervolgens allerlei data verzameld, waarmee de gezondheid van de koe wordt gerapporteerd, maar waarmee ook het (door robots geserveerde) dieet van de koeien automatisch kan worden aangepast.

Ook in de Nederlandse akkerbouw vindt automatisering plaats. Er zijn bijvoorbeeld sorteermachines met computer vision die door machine learning zijn getraind om spruitjes van bepaalde kwaliteitsklassen uit elkaar te houden. Computer vision-systemen die aan trekkers worden bevestigd, kunnen gewassen naar soort en kwaliteit herkennen, om vervolgens per plant zelfstandig een bestrijdings- of bemestingsmiddel toe te dienen, zodat er zo min mogelijk van die kostbare en vaak schadelijke middelen gebruikt wordt. In de kas- en tuinbouw heeft AI de potentie om met computer vision en robotica een deel van het teelwerk over te nemen. De daardoor opgespaarde tijd kunnen telers bijvoorbeeld aan hun bedrijfsstrategie besteden. Ook daarin kan AI een rol spelen. Een grote Nederlandse teler heeft bijvoorbeeld een beeld-herkenningssysteem ontwikkeld dat uit Instagramfoto's van gerechten kan opmaken welke van zijn gewassen het meest in trek zijn.

Ook kan computer vision de kwaliteit van medisch beeldmateriaal verbeteren en chirurgen assisteren bij het uitvoeren van operaties. De Amerikaanse toezichthouder op medische hulpmiddelen (FDA) heeft inmiddels tien diagnosehulpmiddelen met computer vision goedgekeurd voor gebruik in ziekenhuizen.[137] In Nederland is de app SkinVision een voorbeeld van computer vision in de gezondheidssector. Met deze app, die vergoed wordt door zorgverzekeraar CZ, kunnen mensen zelf plekjes op hun huid scannen waarna de app advies geeft over eventuele vervolgstappen.[138] Hoewel er vanuit de medische wetenschap kritiek is op de adequaatheid van dit soort apps[139], laat dit voorbeeld zien op welke manier computer vision in de praktijk wordt toegepast.

Natural language processing

Een derde soort AI-toepassingen is gericht op het automatiseren van het *lezen, analyseren en genereren van menselijke taal*. Het doel van *natural language processing* of taalverwerkingssystemen is dat algoritmes onze 'natuurlijke' taal zo goed mogelijk begrijpen, zodat ze taken kunnen uitvoeren waarvoor het noodzakelijk is tekst te interpreteren. Taalverwerkingsalgoritmes ontleden zinnen op verschillende manieren, bijvoorbeeld door letters en woorden te onderscheiden, tekstdelen te labelen, van links naar rechts te lezen en van rechts naar links. Op die manier kunnen inferenties worden gemaakt over de betekenis van de tekst. Net als computer vision heeft de ontwikkeling van *natural language processing* de afgelopen jaren een impuls gekregen dankzij deep learning. Daardoor is het sneller en gemakkelijker geworden om deze modellen te trainen op menselijke taal.

Omdat taal een sleutelrol speelt in onze communicatie en de manier waarop wij kennis vergaren, bewaren en overdragen, zijn de mogelijkheden van geavanceerde *natural language processing* verstrekkend. Evenals voor computer vision geldt hier dat de huidige toepassingen zich vooralsnog beperken tot specifieke taken waarvoor betrekkelijk weinig wezenlijk begrip van de tekst nodig is. Te denken valt hierbij aan automatische correcties, checks of woordaanvullingen bij het typen van een tekst, maar ook aan het geautomatiseerd vertalen van teksten zoals bij Google Translate.[140] Ook spamfilters en zoekmachines maken gebruik van *natural language processing*. Het algoritme van Googles zoekmachine bijvoorbeeld hanteert voor elke zoekopdracht twee technieken. Allereerst koppelt het de woorden van de zoekopdracht aan de relevante woorden in documenten. Vervolgens rangschikt het algoritme de verschillende

137 Topol 2019: 46.
138 SkinVision, z.d.
139 Freeman et al. 2020.
140 Lewis-Kraus, 14 december 2016.

documenten met de gesignaleerde woorden, op basis van vooronderstelde kwaliteit en relevantie, afgaand op het aantal kliks op de pagina – het beroemde 'PageRank'-algoritme. Deze toepassing van *natural language processing* is revolutionair geweest voor de manier waarop we online zoeken naar informatie. Tegelijkertijd geldt dat hierbij nog geen echt begrip van onze taal vereist is.

Een ander voorbeeld van *natural language processing* zijn 'messenger bots', de geautomatiseerde chatfuncties die veel organisaties als hulpdienst aanbieden op hun website. AI helpt die organisaties om op een efficiënte en snelle manier hun klanten te woord te staan. In feite wordt de verwerking van taal hier gecombineerd met een expertsysteem: het algoritme analyseert een vraag en bepaalt op basis van een beslisboom welk antwoord of welke vervolgvraag het meest gepast is. De Nederlandse politie gebruikt een dergelijke chatbot om mensen te helpen bij onlineaangiftes van internetoplichting. De bot checkt of de aangifte volledig is en adviseert op basis van een eerste beoordeling over de vervolgstappen die het best bij de aangifte passen. Ook het Juridisch Loket ontwikkelt een chatbot, *Julo*, die mensen online te woord staat bij juridische vragen over bijvoorbeeld ontslag. Begin 2021 werd *Julo* gepresenteerd aan minister voor Rechtsbescherming Sander Dekker.[141] De chatbot moet burgers op een laagdrempelige manier toegang geven tot de juiste informatie en hen op die manier helpen zoveel mogelijk zelf hun problemen op te lossen.

Speech recognition

Spraakherkenning of *speech recognition* is het domein binnen AI dat zich bezighoudt met de *waarneming, analyse en interpretatie van gesproken menselijke taal*. Bij *speech recognition* worden algoritmes gebruikt om in gesproken taal zinnen en woorden te ontwaren en die om te zetten in een tekstformat. Een vertaling dus van spraak naar tekst (*speech to text*). Deze vorm van *speech recognition* wordt bijvoorbeeld verkend binnen de gezondheidszorg. Zo wordt er binnen het Leids Universitair Medisch Centrum (LUMC) gewerkt aan een AI-systeem dat gesprekken tussen arts en patiënt middels *automatic speech recognition* omzet in geschreven tekst.[142] Vervolgens wordt die tekst met behulp van *natural language processing* geanalyseerd en identificeert het algoritme daarin specifieke klinische gegevens, zoals informatie over de klachten van een patiënt, en maakt het daar een samenvatting van. Het doel van deze gecombineerde toepassing van *speech recognition* en *natural language processing* is uiteindelijk om de registratielast van artsen te verminderen, waardoor ze meer tijd

141 Rijksoverheid, 12 januari 2021.

142 Van Buchem et al. 2021. Deze studie laat zien dat dit soort *digital scribe*-systemen veelbelovend is. Op dit moment zijn deze systemen echter nog nergens geïmplementeerd en dus kunnen er nog geen uitspraken worden gedaan over de waarde ervan in de praktijk.

over houden voor de patiënt.[143] Dezelfde technologie kan ook andersom worden gebruikt: het omzetten van tekst naar spraak (*text to speech*). Denk bijvoorbeeld aan e-boeken die een computer voorleest, maar ook aan spraakcomputers voor mensen die moeilijk of niet kunnen spreken, zoals veel patiënten met ALS.[144]

Bij slimme stemgestuurde assistenten als Siri (Apple), Google Assistent (Google), Cortana (Microsoft) en Alexa (Amazon) worden beide bovenstaande technieken gecombineerd om gesproken communicatie tussen mens en computer mogelijk te maken. Door te reageren op zogenaamde '*wake words*', zoals 'Siri' of 'Alexa', kunnen deze tools allerlei taken uitvoeren zoals onlinezoekopdrachten, het maken van to-do-lijstjes, het afspelen van muziek en het maken van een restaurantreservering. Het enige dat de gebruiker hoeft te doen, is de opdracht hardop uitspreken. Met *speech recognition* wordt die spraak omgezet in tekst, en met *natural language processing* wordt die informatie geïnterpreteerd en vervolgens bepaald welke actie vereist is.

In tegenstelling tot mensen, die eerder konden spreken en luisteren dan schrijven, zijn computers eerder ingericht op het lezen van tekst dan op het luisteren ernaar. Spraakherkenning is aanzienlijk moeilijker voor computers dan het verwerken van geschreven taal vanwege de variabiliteit en ruis in audiostreams van gesproken taal. Het is een hele uitdaging om die ruis te onderscheiden van spraak, en deze vervolgens om te zetten in het soort tekst dat een computer kan verwerken. Maar ook de spraaksignalen zelf zijn niet 'vanzelfsprekend'. Uit het geluid dat klinkt als wij spreken, is namelijk vaak geen duidelijk onderscheid op te maken tussen afzonderlijke woorden. Vergelijk het met de manier waarop een vreemde taal klinkt: als een spraakwaterval van aaneengeschakelde klanken. Wanneer we een taal niet machtig zijn, is het heel moeilijk om uit de gesproken tekst losse woorden af te leiden; een stap die nog vooráf gaat aan het vertalen van die woorden naar de taal die we wel begrijpen.

Het probleem van *speech recognition* verschilt hierdoor fundamenteel van het interpreteren van geschreven tekst of beeld. Anders dan *computer vision* en *natural language processing*, is de input bij spraakherkenning een enkele variabele – namelijk geluidsgolven – die dynamisch verandert door de tijd. De grote uitdaging is hier dus het onderscheiden van de zinnen en de woorden waaruit die zinnen zijn opgebouwd, om dat geheel vervolgens te vertalen naar een taal waarmee het algoritme aan de slag kan.

---

143 Wouda en Hutink 2019.

144 DeepMind werkt momenteel in samenwerking met Google aan een project om een *tekst-to-speech*-programma te ontwikkelen waarmee mensen met een spraakprobleem via de computer toch met hun eigen stem kunnen spreken (Chen et al. 18 december 2019).

Daarnaast wordt ook een deel van de betekenis overgedragen door de wijze waarop een spreker varieert in volume, cadans en toon – de kenmerkende aspecten van de gesproken taal. Voor de interpretatie van spraak volstaat het dus niet om woorden van elkaar te kunnen onderscheiden. Ook de fonetische aspecten moeten worden opgemerkt en geïnterpreteerd om de betekenis van spraak te reconstrueren. Dan is er nog een ander probleem: homofonen. Oftewel, woorden die hetzelfde klinken maar iets anders betekenen, zoals 'hard' en 'hart' of 'zei' en 'zij'. Om daarbij de juiste betekenis te kunnen bepalen is de context essentieel: wat is de smalle context van de zin, wat is de brede context van de situatie, wie is de spreker, enzovoort.

Doorbraken in machine learning hebben ook op het gebied van spraakherkenning geleid tot vooruitgang doordat grotere hoeveelheden spraakdata kunnen worden verwerkt bij het trainen van deze algoritmes. Het omzetten van spraak naar tekst en tekst naar spraak kan daardoor relatief succesvol worden toegepast in de praktijk, mits er duidelijk – zowel in toon als inhoud – gesproken wordt. Omdat dat bij veel van onze gesproken communicatie niet het geval is, is spraakherkenning nog niet zo geavanceerd dat het op grote schaal voor meerdere taken betrouwbaar kan worden ingezet. Voor een groot deel heeft dat te maken met beperkingen op het gebied van *natural language processing* en het daadwerkelijk begrijpen van taal. Dat AI de stap van lab naar samenleving heeft gemaakt, wil dus allerminst zeggen dat de technologie 'af' is. In de laatste paragraaf van dit hoofdstuk staan we daar uitgebreider bij stil.

Robotica

Met robotica verwijzen we hier naar het soort AI dat wordt toegepast in combinatie met robots. In de robotica komen in wezen alle AI-soorten samen: het vermogen om te redeneren en te leren, om te zien en te horen, om te communiceren en te begrijpen. De robotica onderscheidt zich bij uitstek van andere AI-disciplines door daar nog een fysieke functie aan toe te voegen: *het vermogen om objecten te manipuleren*.

Een robot moet kunnen bewegen en handelen in de fysieke wereld om bepaalde taken uit te voeren. Dat kunnen taken zijn die te eentonig, onprettig of ongezond, gevaarlijk of duur zijn om door mensen te worden uitgevoerd – de *'dull, dirty, dangerous and dear jobs'* –, of waar robots beter in zijn dan mensen. Denk hierbij aan ruimte-expedities, werkzaamheden na de kernramp bij Fukushima en het onklaar maken van bommen.[145] Maar de robotica speelt ook een belangrijke rol bij innovaties binnen de zorg, retail, industrie, veehouderij, land- en

145 Zoals de *PackBot* van het bedrijf Endeavor Robotics (voorheen iRobot), dat onderdeel is van FLIR UGS (*unmanned ground systems*) en robotsystemen levert aan het Amerikaanse leger (FLIR UGS, z.d.).

tuinbouw. In zekere zin zijn ook zelfrijdende auto's als een vorm van robotica te beschouwen. Robots zijn er dus in ontelbare vormen en maten – een uitputtende definitie is dan ook moeilijk te geven. Joseph Engelberger, een pionier op het gebied van industriële robotica, lost het op met een variant op een klassieke oneliner voor ondefinieerbare maar welbekende zaken: *"I can't define a robot, but I know one when I see one."*

Binnen de (klassieke) robotica spelen expertsystemen een belangrijke rol. Dat soort systemen kan worden toegepast voor gestandaardiseerde taken waarbij van tevoren per situatie de gewenste handelingsoptie kan worden bepaald. Niet voor niets worden robots nu met name ingezet in de gecontroleerde omgevingen van fabrieken en havens. Als we robots willen inzetten in de hoog dynamische en vaak chaotische omgeving van ons dagelijks leven, zoals het verkeer, brengt dat complexe uitdagingen met zich mee. Om met zulke variëteit en spontaniteit om te gaan, is een bepaald begrip nodig van hoe dingen in de wereld werken. Robots moeten in staat zijn hun omgeving waar te nemen, situaties te beoordelen, plausibele toekomstscenario's te voorspellen en in een dynamische situatie kunnen bepalen welke van alle mogelijke acties op dat moment gewenst is.[146] Dat soort kennis is de basis voor een systeem dat flexibel genoeg is om zich te verhouden tot de wereld buiten het lab.

Juist omdat robots moeten kunnen omgaan met het '*open ended*'-karakter van onze wereld, waarin de mogelijkheden oneindig zijn, is de integratie van verschillende vermogens cruciaal. Op dit moment vormen de beperkingen van de andere vormen van AI de achilleshiel van de robotica. Robots hebben bijvoorbeeld nog steeds moeite om zelfstandig een donkere handdoek op een donkere tafel te pakken omdat computer vision nog onvoldoende gevoelig is voor licht. Vooruitgang in de andere gebieden van AI en doorbraken in de methoden van machine learning zullen daarom ook cruciaal zijn voor verdere ontwikkelingen in de robotica. De hardware van robotica en de menselijke besturing ervan zijn al vrij indrukwekkend. Toch blijken veel dagelijkse handelingen uiterst verfijnde motoriek, planning en perceptuele vermogens te vergen. Ondanks de ogenschijnlijke overzichtelijkheid van bijvoorbeeld een tuinbouwkas is het voor een robot bijzonder ingewikkeld om een tomaatje te plukken zonder die fijn te knijpen.

Drie grote partijen op het gebied van robotica zijn Boston Dynamics, gespecialiseerd in de simulatie van menselijke beweging in robots ('*humanoids*'), DJI, gespecialiseerd in drones voor consumentengebruik, en Amazon Robotics, gespecialiseerd in geautomatiseerde logistiek. Amazon ontwikkelt en gebruikt

146 Marcus en Davis 2019: 113.

robots die zich efficiënt kunnen bewegen door grote magazijnen en zo sorteerprocessen kunnen optimaliseren. De robots moeten daarvoor rekening met elkaar houden en samenwerken, en doen dat in de sorteercentra heel succesvol. De sorteerrobots hebben specifieke taken en opereren in een – althans, voor robots – voorspelbare en overzichtelijke omgeving.

Boston Dynamics richt zich daarentegen op de ontwikkeling van veel flexibelere robots, zowel fysiek als 'mentaal', zodat ze kunnen worden ingezet voor diverse doeleinden. Het bedrijf, dat in 2017 door Alphabet is verkocht aan het Japanse technologiebedrijf SoftBank, is bekend van de indrukwekkende filmpjes van de twee- en vierbenige robots, zoals Atlas en BigDog. Deze robots kunnen zich voortbewegen en oprichten op manieren die sterk lijken op hoe mensen en dieren zich bewegen. Tot dusver heeft het bedrijf echter nog geen commerciële producten ontwikkeld. DJI, een Chinees technologiebedrijf, doet dat wel. Het bedrijf is wereldmarktleider in onbemande luchtvaartuigen (drones) waarmee gebruikers vanuit de lucht foto's en video's kunnen maken.

In de vorm van robotica, speech recognition, natural language processing, computer vision en machine learning treedt AI zo de samenleving binnen. We hebben dat geïllustreerd aan de hand van voorbeelden van AI-toepassingen in de praktijk. Het is onmogelijk om daarvan een volledig overzicht te geven. In de bijlage van dit rapport geven we voor een aantal sectoren voorbeelden van AI-toepassingen, vooral om een idee te geven van de enorme diversiteit en breedte daarvan. Precies het gegeven dat AI op allerlei manier en plekken wordt ingezet, is illustratief voor AI die breed voet aan de grond krijgt in de samenleving. Die ontmoeting maakt iets los. In de volgende paragraaf gaan we daarom in op de maatschappelijke dynamiek die op gang is gekomen met de stap van lab naar samenleving.

Kernpunten – AI in de praktijk

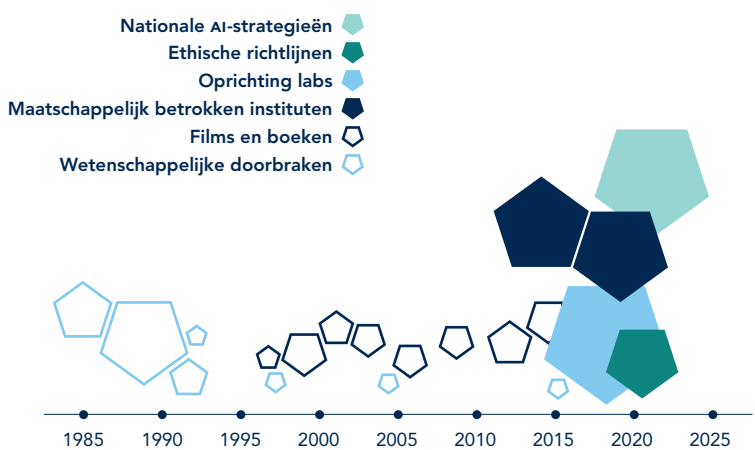
- Met de overstap van lab naar samenleving is AI te vinden in allerlei toepassingen in de praktijk. Daarbij onderscheiden we grofweg vijf soorten AI.
- *Machine learning:* dit soort AI draait om het doen van voorspellende analyses. Een bekend voorbeeld hiervan zijn zogenoemde *recommender systems* die het aanbod van content personaliseren.
- *Computer vision:* hierbij gaat het om het waarnemen en analyseren van visuele informatie, zoals het herkennen van gezichten of bewegwijzering.

- *Natural language processing*: AI gericht op het interpreteren van taal, zoals we die in het dagelijks leven gebruiken. Chatbots werken hier bijvoorbeeld mee.
- *Speech recognition*: dit type AI verwerkt gesproken taal. Spraakassistenten als Siri (Apple) en Alexa (Amazon) maken hier gebruik van.
- *Robotica:* hier komen verschillende vermogens samen, die worden gecombineerd met fysieke functies. Denk bijvoorbeeld aan het vervoeren van goederen in een warenhuis.

2.3 AI als maatschappelijk fenomeen

De overgang van AI als iets dat in het lab wordt onderzocht naar iets dat in de praktijk wordt ingezet, heeft nadrukkelijk ook een maatschappelijke kant. De stap het lab *uit* betekende immers een stap de samenleving *in*. Daar komt AI terecht in een wereld van verschillende krachten en belangen. Er wordt geïnvesteerd en geëxperimenteerd, gediscussieerd en gewaarschuwd. Er verschijnen visies en strategieën om gas te geven met AI, maar ook open brieven en adviezen die juist op de rem trappen. Kortom, deze nieuwe technologie maakt iets los in de samenleving. AI evolueert niet alleen in technologisch opzicht, maar maakt ook als maatschappelijk fenomeen een ontwikkeling door (zie figuur 2.3). Hoewel dit proces nog volop in gang is, kunnen we terugkijkend een aantal tendensen ontwaren. Om beter te begrijpen op welk punt we nu staan, bespreken we hier de verschillende reacties vanuit de samenleving op de komst van AI en de accentverschuivingen die daarbinnen hebben plaatsgevonden.

Figuur 2.3 De ontwikkeling van AI als maatschappelijk fenomeen aan de hand van enkele indicatoren



Aandacht voor AI als revolutionaire technologie

De wetenschappelijke doorbraken die AI deden opbloeien na de meest recente AI-winter, vestigden de aandacht op de talloze mogelijkheden die nu in het verschiet lagen. Er volgden enkele iconische boeken over de toekomst van AI. De nieuwste vorderingen op het gebied van AI worden daarin vaak voorgesteld als het begin van een nieuw tijdperk. Visionair Ray Kurweil speculeert over een aanstaande 'singulariteit', waarin menselijke intelligentie en computerintelligentie zullen versmelten tot een superintelligente eenheid.[147] Wetenschappers Erik Brynjolfsson en Andrew McAfee plaatsen AI in het hart van 'het tweede machinetijdperk', waarin machines naast fysieke taken ook cognitieve taken van de mens overnemen.[148] Filosoof Nick Bostrom ziet daarin een grote dreiging: AI die slimmer en sneller is dan wijzelf, kunnen we steeds moeilijker onder controle houden.[149] Een andere filosoof, Luciano Floridi, spreekt van een 'vierde revolutie' waarin digitale technologieën als AI ons wereldbeeld en zelfbegrip ingrijpend veranderen.[150] Klaus Schwab, oprichter en voorzitter van het World Economic Forum, heeft het over een 'vierde industriële revolutie' waarin de toepassing van slimme technologie onze manier van werken en leven zal transformeren zoals de stoommachine, elektriciteit en digitalisering dat eerder deden.[151]

Een toekomst waarin ons leven nauw verweven raakt met AI, is ook in Hollywood een populair thema. Parallel aan de recente AI-lente verbeelden films als Her (2013), Ex Machina (2014) en Transcendence (2014) een toekomst waarin AI een cruciaal intelligentieniveau bereikt. Ze fungeren daarmee in feite als een vorm van *scenario thinking*: ze tonen scenario's waarin we ons ook emotioneel gaan verhouden tot AI en er zelfs verliefd op kunnen worden (*Her*),waarin AI de ultieme Turingtest doorstaat en zo mensgetrouw wordt dat we mens en machine niet meer kunnen onderscheiden (*Ex Machina*), of zich ontwikkelt tot een gevaarlijke, haast niet te controleren bron van macht (*Transcendence*). Hoewel de thematiek van superintelligente computers al langer inspiratie biedt voor filmscripts, maken deze laatste producties specifiek de term 'AI' bekend bij het grote publiek.

Toegepast onderzoek en de run op talent

Naast deze meeslepende beeltenissen waren het vooral ook de praktische mogelijkheden van nieuwe AI-technieken waardoor het onderwerp op de niet-academische agenda's kwam te staan. Die toepasbaarheid maakte AI

147 Kurzweil 2005.
148 Brynjolfsson en McAfee 2014.
149 Bostrom 2016.
150 Floridi 2014.
151 Schwab 2016.

vanuit economisch oogpunt namelijk interessant voor zowel het bedrijfsleven als overheden. We zagen eerder dat de private investeringen in AI wereldwijd aanzienlijk zijn toegenomen. Bedrijven slaan bovendien de handen ineen met kennisinstellingen in speciale 'AI-labs', waar ze fundamenteel onderzoek verbinden aan behoeften uit de praktijk.

In 2015 werd het eerste praktijkgerichte AI-lab in Nederland gelanceerd: het QUVA Deep Vision Lab, een samenwerking tussen de Universiteit van Amsterdam en Qualcomm waarin onderzoek op het gebied van computer vision wordt vertaald naar industriële toepassingen. In de jaren daarna volgen in rap tempo meer van dit soort samenwerkingen waarin de stap van wetenschappelijke inzichten naar innovatieve toepassingen in de praktijk centraal staat. Het Innovation Center for Artificial Intelligence (ICAI), opgericht in 2018 door de Universiteit van Amsterdam en de Vrije Universiteit, speelt daarbij een belangrijke coördinerende en ondersteunende rol. Inmiddels telt ons land twintig ICAI-labs waarin bedrijven als Bosch, TomTom, KPN, ING, Ahold-Delhaize en DSM, maar ook ziekenhuizen, de nationale politie en overheden samenwerken met universiteiten en onderzoeksinstituten om tot innovatieve AI-oplossingen te komen voor vraagstukken uit de praktijk. De toepassingsgebieden lopen uiteen van landbouw en mobiliteit tot zorg en retail. Tabel 2.1 geeft een overzicht.

Tabel 2.1 Overzicht van AI-labs in Nederland

Lab	Samenwerking	Focus	Domein
AI for Agro-Food Lab	Wageningen University & Research, OnePlanet, 4TU, Avular, Connecterra, Imec, Provincie Flevoland, Kubota, Leafteasers, Lely, NanoPHAB, NXP Semiconductors, Sensor Sense, Settels Savenije, Signify	Robotica, machine learning, deep learning, computer vision	Landbouw en voedsel
AI for Biosciences Lab	DSM, TU Delft	Machine learning	Bioproductie
AI for Fintech	ING, TU Delft	Machine learning	Financiële sector
AI for Precision Health, Nutrition & Behavior	Radboud Universiteit, Radboud UMC, OnePlanet, Wageningen University & Research, Artinis Medical Systems, ConnectedCare, Flow.ai, Imec, Noldus, Orikami, Mead Johnson Nutrition, Salut, Soa Aids Nederland	Machine learning, natural language processing, robotica	Gezondheidszorg (preventief)
AIM Lab	Inception Institute of AI, Universiteit van Amsterdam	Computer vision	Gezondheidszorg (medical imaging)

Lab	Samenwerking	Focus	Domein
AIR Lab	Ahold-Delhaize, TU Delft, Universiteit van Amsterdam	Machine learning	Retail
Atlas Lab	TomTom, Universiteit van Amsterdam	Computer vision	Navigatie
Civic AI Lab	Gemeente Amsterdam, Universiteit van Amsterdam, Vrije Universiteit	Machine learning	Publieke sector
Cultural AI Lab	Centrum voor Wiskunde en Informatica, KNAW, Koninklijke Bibliotheek, Rijksmuseum, Nederlands Instituut voor Beeld en Geluid, TNO, Universiteit van Amsterdam, Vrije Universiteit	Symbolische AI, machine learning	Culturele sector
DELTA Lab	Bosch, Universiteit van Amsterdam	Computer vision, deep learning	Mobiliteit
Discovery Lab	Elsevier, Universiteit van Amsterdam, Vrije Universiteit	Machine learning	Wetenschappelijk onderzoek
Donders AI for Neurotech Lab	Donders Instituut, Radboud Universiteit, Phosphoenix, Advanced Bionics, OnePlanet, Abbott	Machine learning	Neurotechnologie
EAISI AI-enabled Manufacturing & Maintenance (AIMM) Lab	TU Eindhoven, KMWE, Lely, Marel, Nexperia	Machine learning, computer vision, robotica	Industriële toepassing
EAISI fast Lab	TU Eindhoven, Lely, Rademaker, Diversey, ExRobotics, Vanderlande	Robotica	Industriële toepassing
EAISI Mobility Lab	TU Eindhoven, NXP Semiconductors	Computer vision, machine learning	Mobiliteit
KPN Responsible AI Lab	KPN, Jheronimus Academy of Data Science	Deep learning, machine learning, natural language processing	Telecom
Police Lab AI	Nationale politie, Universiteit Utrecht, Universiteit van Amsterdam	Machine learning, symbolische AI	Veiligheid
QUVA Lab	Qualcomm, Universiteit van Amsterdam	Computer vision, deep learning	Industriële toepassing
Radboud AI for Health	Radboud Universiteit, Radboud UMC, JADS	Machine learning, deep learning	Gezondheidszorg
Thira Lab	Thirona, Delft Imaging Systems, Radboud UMC	Deep learning	Gezondheidszorg (medical imaging)

Met bedrijven die de mogelijkheden van AI op allerlei terreinen verkennen en toepassen, is bovendien een enorme vraag ontstaan naar AI-talent, en een discussie over de capaciteit om dat talent in ons land te ontwikkelen en te behouden. In Nederland zijn de afgelopen drie jaar meer dan tien nieuwe hoogleraren benoemd op het gebied van AI.[152] Aan het begin van dit hoofdstuk lieten we zien dat de meerderheid van de promovendi in de VS voor een commerciële baan kiest. Ook Nederlandse hoogleraren waarschuwen voor een 'braindrain' binnen het vakgebied naar het bedrijfsleven en het buitenland.[153] Volgens hoogleraar Intelligent Sensory Information Systems Cees Snoek (UvA) vertrekken toponderzoekers vaak naar grote techbedrijven.[154] Hoogleraar Machine learning Holger Hoos (UL) zegt in twintig jaar AI-onderzoek 'niet deze gekte' te hebben gezien en ook Maarten de Rijke, hoogleraar AI & Information retrieval (UvA), luidt publiekelijk de noodklok over een reële talentvlucht.[155] Ze laten daarmee een belangrijk geluid horen in de aandacht voor de kansen die AI te bieden heeft: "zonder talent kun je het schudden".[156]

Actieplannen voor AI

De gerichtheid op de mogelijkheden die AI biedt, is ook terug te zien in nationale AI-strategieën. Over het algemeen staan hierin de economische kansen van AI centraal. Ook de OESO stelt vast dat deze nationale strategieën veelal tot doel hebben om met AI de nationale productiviteit en competitiviteit te vergroten.[157] De actieplannen zijn daarom in eerste instantie vaak gericht op de ontwikkeling van AI, onder andere door onderzoek te stimuleren, te werken aan ondersteunende infrastructuren en het bedrijfsleven te stimuleren. In een groot deel van deze nationale strategieën gaat eveneens aandacht uit naar de maatschappelijke en ethische dimensie van AI. Vaak volgen deze passages echter na de uiteenzetting van de economische plannen en zijn ze over het algemeen minder concreet en actiegericht. De argumentatie achter dat verschil in nadruk is in Europa doorgaans dat het alleen mogelijk is om AI vanuit onze waarden richting te geven als we technologisch gezien meedoen met de voorhoede.

Om die reden dringt denktank DenkWerk in 2018 in het rapport *AI in Nederland* aan op de urgentie voor Nederland om serieus aan de slag te gaan met AI. In het rapport stelt de denktank dat Nederland internationaal achterloopt als het

---

152 Onder andere Cees Snoek (UvA), Maarten de Rijke (UvA), Antal van den Bosch (UvA), Natali Helberger (UvA), Ivana Isgum (UvA), Tobias Blanke (UvA), Koen Hindriks (VU), Jan Broersen (UU), Mehdi Dastani (UU), Mark Winands (UM), Tibor Bosse (RU).

153 Onder andere Van Noort, 27 augustus 2018.

154 NOS, 1 oktober 2018.

155 Heest, 3 maart 2020.

156 Heest, 3 maart 2020.

157 OESO 2019: 121.

gaat om investeringen in de private sector en andere ondersteuning vanuit de overheid. Volgens DenkWerk werd het gesprek over AI destijds vooral op toon van ongerustheid gevoerd, waardoor de kansen die de technologie te bieden heeft uit het oog dreigen te raken. De denktank wijst op het 'enorme maatschappelijke potentieel' van AI en roept de overheid op om tot een nationale agenda te komen voor de ontwikkeling en toepassing van AI. Ze maant daarbij tot actie: "dit is geen dossier om eerst nog eens twee jaar op te gaan studeren."[158]

In datzelfde jaar werkt DenkWerk mee aan een aanzet tot een nationale AI-agenda. AINED, een samenwerking tussen TopTeam ICT, VNO-NCW, ICAI, NWO, TNO en ondersteund door BCG, publiceert een rapport waarin de door DenkWerk benadrukte urgentie gestalte krijgt. In het document concluderen de betrokken partijen dat Nederland de mogelijkheden van AI onvoldoende benut en internationaal achterop raakt. AINED roept op om van AI een nationale prioriteit te maken om zo de welvaart en de internationale positie van Nederland veilig te stellen en te versterken. Om de ontwikkeling van AI te versnellen en ons land internationaal te differentiëren formuleert AINED enkele doelen die het uitgangspunt zouden moeten zijn bij een nationale strategie.

Die strategie komt er uiteindelijk in het najaar van 2019. Uit het AINED-rapport vloeit een task force voort onder leiding van VNO-NCW en het ministerie van Economische Zaken en Klimaat (EZK) die vervolg moet geven aan de geformuleerde doelen. De eerste grote concrete actie is het oprichten van een Nederlandse AI-coalitie, een platform van bedrijven, overheden, maatschappelijke organisaties, en kennis- en onderzoeksinstellingen die zich gezamenlijk inzetten als katalysator van AI-ontwikkeling in Nederland. Deze Nederlandse AI Coalitie (NL AIC) wordt gelanceerd in 2019 tijdens de conferentie *Nederland Digitaal*. De coalitie kondigt daar aan het opzetten van AI-labs te willen stimuleren en samen met het kabinet een AI-strategie te gaan ontwikkelen.

In oktober 2019 presenteert EZK, mede namens de ministeries van Justitie en Veiligheid (JenV), BZK, Sociale Zaken en Werkgelegenheid (SZW) en Onderwijs, Cultuur en Wetenschap (OCW), het eerdergenoemde Strategisch Actieplan voor AI (SAPAI). Het kabinet beschrijft hierin de koers voor de aankomende jaren, met de eerste concrete acties om de AI-ontwikkeling en de profilering van Nederland op dit gebied te versnellen. SAPAI volgt een driesporenbeleid. Spoor 1 is gericht op het benutten van de maatschappelijke en economische kansen die AI biedt, spoor 2 op het creëren van het juiste ecosysteem daarvoor en spoor 3 op de waarborgen om dit alles op een verantwoorde manier te realiseren. Dit laatste reflecteert een verschuiving die zich inmiddels in het debat over AI

158 Denkwerk 2018.

voordoet, waarbij naast de economische kansen ook steeds meer aandacht ontstaat voor de effecten van AI-toepassingen. Het SAPAI wordt dan ook gelijktijdig aangeboden met de Kamerbrieven 'AI, publieke waarden en mensenrechten'[159] en 'Waarborgen tegen risico's van data-analyses door de overheid'[160].

De private investeringen, de oprichting van AI-labs en de lancering van nationale AI-strategieën kunnen als indicatief worden gezien voor de groeiende aandacht voor AI buiten de wetenschap. Het accent ligt dan op de nieuwe mogelijkheden die AI biedt. Het bewustzijn daalt in dat AI een zekere jongvolwassenheid heeft bereikt en dat verschillende partijen in de samenleving aan zet zijn om de potentie ervan te realiseren. AI komt op de agenda te staan, en dan met name op de economische agenda. Media berichten wekelijks over nieuwe toepassingen en de wereld die daarmee wordt geopend. Op die manier wordt ook het bredere publiek zich bewust van het fenomeen. Velen weten dan nog niet wat AI precies is, maar wel dát het er is.

Aandacht voor de effecten van AI in de praktijk

In het kielzog van de brede aandacht die AI trekt, ontwikkelt zich de vraag naar de impact ervan. De introductie van AI in de echte wereld is aanleiding om na te denken over de implicaties daarvan voor ons dagelijks leven. In de nieuwe boeken die verschijnen, verschuift het accent van het revolutionaire karakter van AI naar de consequenties die de toepassing van AI kan hebben in het echte leven. Verschillende auteurs schrijven over de problemen die zich kunnen voordoen bij de stap die AI maakt van het lab naar de wereld waarin wij leven. Ze wijzen op de morele, sociale, politieke, juridische en economische vraagstukken die zich daarbij voordoen. Er ontstaat dus een discussie over het effect dat AI heeft op onze waarden.

Met haar boek *Weapons of Math Destruction* (2017) waarschuwt Cathy O'Neil voor de destructieve effecten die een onvoorzichtig en kortzichtig gebruik van algoritmes kan hebben op het leven van mensen.[161] Meredith Broussard verkondigt eenzelfde boodschap en wijst erop hoe 'technochauvinisme', het idee dat technologie geschikt is om elke vraag te beantwoorden, mensen heimelijk tekort kan doen.[162] Shoshana Zuboff richt zich niet zozeer op de technologie maar vooral op de spelers daarachter. Ze slaat alarm voor wat zij 'surveillancekapitalisme' noemt, een economisch logica van excessieve dataverzameling en voorspellende algoritmes waarmee grote techbedrijven op ongekende wijze

159 Kamerstukken II 2019/20, 26 643, nr. 642.
160 Kamerstukken II 2019/20, 26 643, nr. 641.
161 O'Neil 2017.
162 Broussard 2019.

invloed kunnen uitoefenen op ons gedrag. In *The Algorithmic Society* (2020) wijzen verschillende auteurs eveneens op de relatie tussen data, algoritmes en macht, en ze beschrijven hoe dat de verhouding tussen de staat en burgers kan veranderen. Ook Stuart Russell, auteur van het leidende AI-handboek, wijdt zijn laatste boek aan de effecten van AI in de echte wereld. Zijn boodschap: systemen kunnen technisch goed werken, maar toch ongewenste gevolgen hebben, en dus is het zaak om ten alle tijden controle te houden over AI: *"What's worse than a society-destroying AI? A society-destroying AI that won't switch off."*[163]

Samenleven met AI wordt een belangrijk thema. In 2016 staat het World Economic Forum in het teken van het vormgeven van een wereld met slimme technologieën als AI. In datzelfde jaar spreken G7-ministers op het gebied van ICT samen met de OESO af om een internationale discussie te starten over de ontwikkeling van AI en de economische en sociale implicaties daarvan. De OESO gaat zich vanaf dan steeds actiever bezighouden met AI, organiseert conferenties en stimuleert internationale beleidsdiscussies. In 2019 presenteert de OESO haar principes voor AI, die aandacht voor de uitwerking van AI op mens en maatschappij thematiseren en richting moeten geven aan een verantwoorde ontwikkeling. Alle OESO-lidstaten, de G20 en een tiental andere landen nemen de principes aan, die daarmee de eerste intergouvernementele richtlijnen voor AI vormen.

Maatschappelijke organisaties raken betrokken

De aandacht voor de economische impact van AI verschuift naar aandacht voor de maatschappelijke impact ervan. Op internationaal niveau gaat ook UNESCO zich daarmee bezighouden. In 2018 wijdt de organisatie haar magazine *The UNESCO Courier* geheel aan de kansen en bedreigingen voor de samenleving. Directeur-generaal Audrey Azoulay wijst in die editie op het belang van een ethisch debat over AI en ziet daarin een rol weggelegd voor UNESCO: "*It is our responsibility (…) to enter this new era with our eyes wide open.*"[164] UNESCO geeft die verantwoordelijkheid vorm door eveneens te werken aan een wereldwijde ethische standaard voor AI die als basis moet dienen voor de ontwikkeling van nationaal beleid.[165] In Europa heeft de Europese Commissie dan inmiddels de High-Level Expert Group on AI (AI HLEG) opgericht, het comité dat Europese overheden moet voorzien van een ethisch kader voor de technologie. Na de lancering van de Europese strategie voor de ontwikkeling van AI staat nu ook de *verantwoorde* ontwikkeling op de agenda, met aandacht voor de effecten daarvan. De richtlijnen en aanbevelingen die de AI HLEG presenteert voor

163 Sample, 24 oktober 2019.
164 Šopova 2018.
165 In 2020 is de draft van dit document gepresenteerd (Ad Hoc Expert Group, 7 september 2020).

'vertrouwenswaardige AI' moeten beleidsmakers bewust maken van de ethische en maatschappelijke aspecten van AI en handvatten bieden om daarmee om te gaan.[166]

Termen als 'mensgerichte AI' (*human-centric AI*), ethische, humane en verantwoorde AI duiken steeds vaker op en indiceren een toenemende aandacht voor de manier waarop de inzet van AI zich verhoudt tot een wereld van mensen en waarden. Er worden onderzoeksinstituten opgericht die zich specifiek richten op de maatschappelijke implicaties van AI. Zelfs in Silicon Valley, het hart van de technologische ontwikkeling van AI, wordt een onderzoeksinstituut opgericht dat zich specifiek richt op de menselijke en maatschappelijke impact van AI: het Stanford Institute for Human-Centered AI.

Het AI Now Institute, eveneens uit de VS, is misschien wel het meest prominente voorbeeld van een onderzoeksinstituut dat zich heeft toegelegd op de sociale en maatschappelijke effecten van AI. Het is met zijn jaarlijkse rapporten wereldwijd een belangrijke aanjager van dat debat. Sinds het eerste rapport uit 2016 is AI Now zich steeds duidelijker gaan uitspreken: waar de aanbevelingen aanvankelijk aanstuurden op het in kaart brengen van de effecten van AI, pleit het instituut in latere rapporten voor een (voorlopig) verbod op bepaalde toepassingen en stelt het specifieke eisen aan een verantwoorde inzet van AI.

De veranderde toon van deze aanbevelingen is illustratief voor de ontwikkeling die het maatschappelijk debat over AI de afgelopen jaren heeft doorgemaakt: van bewustwording van de effecten van AI naar een inhoudelijke discussie over hoe bepaalde waarden kunnen worden geraakt en beschermd. In Nederland levert in eerste instantie het Rathenau Instituut een belangrijke bijdrage aan die discussie. Met het rapport *Opwaarderen: Borgen van publieke waarden in een digitale samenleving* (2017) agendeert het Rathenau Instituut de brede maatschappelijke impact van slimme technologieën als AI. De hoofdboodschap van het rapport is dat de Nederlandse samenleving niet klaar is voor de nieuwe digitaliseringsgolf en dat publieke waarden, zoals gelijke behandeling, menselijke waardigheid en ongelijke machtsverhoudingen, daardoor onder druk kunnen komen te staan.[167] Het Rathenau Instituut verbreedt daarmee de discussie die tot dan toe voornamelijk in het teken van privacy en veiligheid stond.

'*Opwaarderen*' heeft een doorslaggevende agenderende functie gehad en binnen de overheid aandacht gegenereerd voor de maatschappelijke impact van nieuwe digitale technologieën. De aanbevelingen waren dan ook voornamelijk

166 High-Level Expert Group on Artificial Intelligence 2019b en 2019c.
167 Kool et al. 2017.

gericht op het opstarten van een politieke en maatschappelijke dialoog over deze impact. In de jaren daarop is het Rathenau Instituut zich gaan richten op specifieke technologieën als virtual reality, surveillancetechnologie en sensoren, met vaak een belangrijke rol voor AI. De aanbevelingen aan het kabinet worden dan zowel operationeler als uitgesprokener: het accent verschuift van descriptieve naar prescriptieve analyses. Door ervaring en onderzoek wordt duidelijker hoe nieuwe technologieën in de praktijk kunnen uitwerken, en daardoor worden de ideeën over wat er nodig is om onwenselijke effecten tegen te gaan ook concreter. Het debat waar enkele jaren geleden voor werd gepleit, begint zo steeds meer invulling te krijgen.

Ook onderzoeksinstituut Waag levert daaraan in Nederland een belangrijke bijdrage, in het bijzonder bij monde van oprichter en directeur Marleen Stikker. Zij vraagt aandacht voor de vraagstukken waar 'echte AI', dus niet de speculatieve AI van de toekomst, ons nu al mee confronteert en benadrukt de rol die het ontwerp van de technologie daarbij speelt. In hoofdstuk 6 staan we uitgebreider stil bij het werk van Stikker en Waag.

Onderzoek en advies

Naarmate meer zicht ontstaat op hoe AI in de praktijk kan worden toegepast gaan ook andere onderzoeks- en adviesbureaus die niet primair georiënteerd zijn op technologie, zich het thema aantrekken. Waar AI aanvankelijk nog als onderdeel van het bredere thema van digitalisering werd besproken[168], richten organisaties op het gebied van onder andere onderwijs, zorg, veiligheid, infrastructuur en recht zich de laatste jaren specifiek op de vraag wat de komst van AI betekent voor hun veld. Met hun onderzoek naar autonome wapensystemen zijn de AIV en de Commissie van advies inzake volkenrechtelijke vraagstukken (CAVV) een van de eerste instituten in Nederland die de impact van AI binnen een specifiek domein onderzochten.[169] In hun rapport agenderen zij de militaire toepassing van AI en de noodzaak van betekenisvolle menselijke controle over dit soort geautomatiseerde wapens. Het Planbureau voor de Leefomgeving (PBL) bracht in kaart wat de komst van slimme algoritmes betekent voor de publieke waarden die centraal staan in het mobiliteitsdomein.[170] Met behulp van AI kunnen we volgens het PBL onze infrastructuur beter benutten, maar zulke toepassingen kunnen ook de toegankelijkheid en veiligheid daarvan onder druk zetten en vragen om nieuwe spelregels.

168 Onder andere Raad van State 2018; Onderwijsraad 2017; SER 2016.
169 AIV en CAVV 2015.
170 PBL 2017.

Vanaf 2018 verschijnen in Nederland meer, vaak verkennende sectoradviezen waarin wordt gereflecteerd op de betekenis van AI voor een specifiek domein. Zo onderzocht het Centrum voor Ethiek en Gezondheid (CEG) welke vraagstukken de inzet van AI in de zorgpraktijk oplevert[171] en bracht ook de Raad voor Volksgezondheid en Samenleving (RVS) een verkenning uit over de implicaties van AI-gebruik voor de waarden binnen de zorg.[172] Beide onderzoeken maken duidelijk dat er in de praktijk nog veel werk te doen is om AI binnen de zorgsector goed te laten functioneren. Binnen het onderwijs vinden vergelijkbare verkenningen plaats. Op verzoek van OCW onderzocht Dialogic de impact van AI op het onderwijs in Nederland.[173] Het advies is, net als andere vroege adviezen, voornamelijk gericht op voorbereidende maatregelen, zoals het werken aan acceptatie, vaardigheden en de benodigde infrastructuur, en op leren via experimenten en verder onderzoek.

Met deze verkenningen en adviezen krijgt de AI-discussie in Nederland meer reliëf: de verschillende contexten waarin AI kan worden toegepast, maakt duidelijk hoe breed en veelzijdig de impact van AI zal zijn. Intussen wordt er meer ervaring opgedaan met de inzet van AI en daaruit blijkt vooral welke moeilijkheden en risico's zich voordoen in de stap naar de praktijk. Voorbeelden van discriminatie door algoritmes, ongelukken met zelfrijdende auto's en het verdwijnen van de menselijke maat door een doorgeschoten 'algoritmisering' geven te denken over de manier waarop we AI in de samenleving willen integreren.

De schaduwzijde

Er ontstaan discussies over de rol van AI bij de ontwikkeling van autonome wapens, het gebruik van gezichtsherkenning door steden en politie, en de positie van Big Tech. Het College voor de Rechten van de Mens publiceerde dit jaar een handreiking voor overheden met uitgangspunten voor (semi-)geautomatiseerde besluitvorming, vanuit de zorg dat het gebruik van AI het risico op discriminatie kan vergroten en afbreuk kan doen aan de rechtsbescherming.[174] Ook jurist en wetenschapper Marlies van Eck en filosoof en schrijver Maxim Februari zijn in Nederland belangrijke aanjagers van het publieke debat over het overheidsgebruik van algoritmes. In hoofdstuk 6 gaan we nader in op hun rol in deze discussie.

171 CEG 2018.
172 RVS 2019.
173 Van der Vorst et al. 2019.
174 College voor de Rechten van de Mens 2021.

Voorbeelden van hoe het mis kan gaan met de toepassing van AI in combinatie met de stroomversnelling waar de technologie zich momenteel in bevindt, leggen bovendien een bepaalde druk op de discussie: om te voorkomen dat er onwenselijke situaties ontstaan door het toenemend gebruik van AI, is het niet alleen van belang om zicht te krijgen op de risico's maar ook om daarnaar te handelen. Zo hebben verschillende Amerikaanse staten en steden het gebruik van gezichtsherkenning door de politie of in openbare ruimten verboden.[175] Ook de Europese Commissie stelt in de concept-Verordening voor AI een verbod voor op grootschalige gezichtsherkenning in de publieke ruimte, maar laat daarbij ruimte voor uitzonderingen vanuit bijvoorbeeld veiligheidsoverwegingen – door sommigen bestempeld als 'achterdeurtjes'.[176]

Ook actiegroepen en belangenorganisaties komen in beweging. In Nederland wijzen onder andere Bits of Freedom en Stichting Kafkabrigade op de manier waarop AI onrecht in de hand kan werken en bestaande ongelijkheden kan verdiepen. In de eerste plaats gaat het dan om organisaties die vanuit hun focus op een bepaald type problemen aandacht ontwikkelen voor de misstanden die kunnen ontstaan rondom het gebruik van AI. Gaandeweg ontstaan er ook initiatieven die zich specifiek bezighouden met deze thematiek, zoals het Duitse AlgorithmWatch, dat het internationale gebruik van algoritmische systemen in kaart brengt en kritisch evalueert. Inmiddels staat AI ook op de agenda van de grotere organisaties op het gebied van mensenrechtenbescherming zoals Amnesty International, Hivos, Human Rights Watch en UNICEF.[177] Kortom, er mobiliseert zich een maatschappelijk middenveld dat zich richt op de effecten van AI.

Op de politieke agenda

In de politiek groeit eveneens de aandacht voor de vraagstukken rondom AI. Die vraagstukken zijn eerst onderdeel van de bredere discussie over digitalisering en privacy, zoals we ook zagen bij de onderzoeksinstituten en adviesraden. De digitaliseringsdiscussie staat begin jaren tien van deze eeuw in het teken van Big Data en privacy. In het regeerakkoord uit 2012 leggen PvdA en VVD vast dat de uitvoering van een Privacy Impact Assessment (PIA) verplicht wordt bij de bouw van ICT-systemen en het aanleggen van databestanden. Op dat moment wordt al gewerkt aan de Algemene Verordening Gegevensbescherming (AVG),

175 In de staten Californië, Oregon en New Hampshire is het gebruik van biometrische surveillancetechnologie door de politie verboden, en in de steden Portland, San Francisco en Oakland is gezichtsherkenning in openbare ruimten in het geheel verboden.

176 Europese Commissie 2021b; Verhagen, 21 april 2021.

177 Amnesty heeft sinds 2016 een aparte afdeling, Amnesty Tech, die zich richt op kwesties die te maken hebben met digitale technologie, waaronder AI.

die een wettelijke basis moet bieden om persoonsgegevens, zeker in het digitale domein, beter te beschermen.

In het verlengde van de focus op privacy ligt de aandacht voor transparantie. In de notitie *Vrijheid en veiligheid in de digitale samenleving: een agenda voor de toekomst* uit 2013 staat de impact van digitale technologie op privacy centraal en wordt transparantie als bijzonder belangrijk beginsel gezien bij het beschermen daarvan. Ook in de voorstellen voor de AVG krijgt het transparantiebeginsel dan al een belangrijke rol toebedeeld.[178] Dit zijn ook twee waarden die vanaf het begin dominant zijn in de politieke discussie rondom AI. Naast een algemenere discussie over digitalisering, gaat de aandacht voor Big Data geleidelijk over in een debat over de manier waarop die data worden verwerkt middels steeds slimmer wordende algoritmes. Ook de WRR wijst in het advies *Big Data in een vrije en veilige samenleving* op de vitale rol die algoritmes spelen in Big Data-processen.[179]

Vanaf 2018 ontstaat er, in de vorm van moties, vaker aandacht in de Tweede Kamer voor het gebruik van algoritmes en AI, gevoed door de onderzoeken en adviezen die rond die tijd veelvuldig het licht zien en door de jongste ervaringen met algoritmegebruik in de praktijk. De werking en uitwerking van AI worden als expliciet thema geagendeerd in het politieke debat. De inzet van SyRI en het gebruik van algoritmes bij de Belastingdienst hebben hieraan een impuls gegeven. Dat debat wordt dan met name gedreven door de ontwrichtende effecten van AI en de zorgen over de negatieve consequenties van algoritmegebruik.

Naast de aandacht voor privacyrisico's en het belang van transparantie ontstaat oog voor vraagstukken rondom discriminatie[180] en menselijke controle.[181] In 2020 wijdt de Tweede Kamer voor het eerst een Algemeen Overleg aan AI en sleuteltechnologieën. Daarin reageert de Kamer op een vijftal brieven van de staatsecretaris van EZK.[182] Naast vragen over de Nederlandse investering in deze technologieën worden in dit overleg ook veelvuldig vragen gesteld over de risico's op het gebied van privacy, discriminatie en oncontroleerbare algoritmes. Ook wordt de inzet van AI door specifieke partijen zoals de politie[183] of de overheid[184] ter discussie gesteld.

178 Kamerstukken II 2013/14, 26 643, nr. 298.
179 WRR 2016.
180 Kamerstukken II 2018/19, 32 761, nr. 138; Kamerstukken II 2019/20, 30950, nr. 206.
181 Kamerstukken II 2018/19, 35 212, nr. 2.
182 Kamerstukken II 2019/20, 26 643, nr. 681.
183 Kamerstukken II 2019/20, 26 643, nr. 652.
184 Kamerstukken II 2017/18, 32 761, nr. 117.

Het zijn met name de kortetermijneffecten van AI-toepassingen waar de aandacht naar uitgaat. In reactie daarop wordt dan ook gewerkt aan concrete manieren om de risico's het hoofd te bieden. De ontwikkeling van normen voor AI-toepassingen door de NEN is daarvan een voorbeeld, maar ook de Kamerbrieven over de waarborgen tegen de risico's van data-analyses bij de overheid[185] en het gebruik van gezichtsherkenningstechnologie[186] kunnen worden gezien als acties om de directe risico's van AI-gebruik te adresseren.

Parallel aan de discussies over het effect van AI op specifieke waarden en in specifieke contexten duikt de term 'publieke waarden' vaker op als het over digitalisering gaat. Het is onder meer het Rathenau Instituut dat het kabinet met het rapport *Opwaarderen* uitnodigt om vanuit dat perspectief, met een brede blik op de maatschappelijke impact, na te denken over de opkomst van digitale technologieën als AI. In de Nederlandse Digitaliseringsstrategie (2018) en het latere SAPAI (2019) wordt de borging van publieke waarden en mensenrechten genoemd als fundament voor een samenleving waarin digitale technologie steeds verder verweven raakt.[187]

BZK geeft met de beleidsbrief over AI, publieke waarden en mensenrechten het startschot voor een uitgebreid programma over deze thematiek. Met de aandacht voor de integrale impact van digitale technologie op onze samenleving dient zich ook de vraag aan naar de grip op deze ontwikkelingen vanuit de politiek. In juli 2019 wordt de Tijdelijke Commissie Digitale Toekomst in het leven geroepen om te onderzoeken hoe de Tweede Kamer omgaat met vraagstukken rondom nieuwe digitale technologie. De commissie oordeelt dat er onvoldoende kennis en aandacht is om deze ontwikkelingen in hun volle breedte te overzien en richting te geven.[188] Aanbevolen wordt onder andere om een vaste commissie Digitale Zaken op te stellen om structureel zicht te krijgen op ontwikkelingen op het gebied van digitalisering, een advies dat de Kamer overneemt. Sinds april 2021 kent het parlement een vaste Kamercommissie voor Digitale Zaken (DiZa), die als doel heeft om tegen de achtergrond van nieuwe technologieën als AI "overzicht te creëren en voor verbinding te zorgen in de behandeling van digitale dossiers op de diverse beleidsterreinen".[189]

185 Kamerstukken II 2019/20, 26 643, nr. 641.
186 Kamerstukken II 2019/20, 32 761, nr. 152.
187 EZK 2018.
188 Tijdelijke Commissie Digitale Toekomst 2020.
189 Tweede Kamer, z.d.

Dat AI op de politieke agenda staat, blijkt ook uit de verzoeken van de overheid aan universiteiten en onderzoeksbureaus om studies te verrichten naar de impact van AI. Eerder noemden we al het onderzoek van Dialogic naar de betekenis van AI voor het onderwijs. In opdracht van het WODC onderzocht de Universiteit van Tilburg de risico's van gezichtherkenningstechnologie voor de privacy van burgers.[190] In het rapport identificeren de onderzoekers verschillende privacyrisico's, zoals een aantasting van autonomie, ongelijke machtsverhoudingen en secundair gebruik van data. Op aanvraag van BZK verrichtte de Universiteit Utrecht een systematische, juridische studie naar de manier waarop algoritmes zich verhouden tot de Nederlandse grondrechten. De slotsom van dit onderzoek luidt dat met name gelijke behandelingsrechten en procedurele rechten kunnen worden aangetast door de inzet van slimme algoritmes.[191] Daarmee worden *bias* en transparantie onderstreept als belangrijke thema's in het publieke debat over AI.

In reactie op de risico's van AI die zich beginnen af te tekenen, klinkt steeds vaker de roep om een aparte AI-toezichthouder of 'algoritmewaakhond'. Daarop heeft BZK opdracht gegeven voor een onderzoek naar het toezicht op het gebruik van algoritmes door de overheid. Adviesbureau Hooghiemstra & Partners concludeert dat er op dit moment geen juridische lacunes bestaan en er dus geen aanleiding is om een nieuwe toezichthouder in het leven te roepen. Toezichthoudende instanties beschikken over voldoende bevoegdheden om ook in het kader van algoritmegebruik effectief hun werk te kunnen doen. Deze onderzoeksverzoeken laten zien dat de regering bezig is zich van buitenaf te laten informeren over wat de inzet van AI betekent voor het functioneren van de overheid en haar verantwoordelijkheid om toe te zien op de bescherming van rechten en andere publieke waarden.

Ethiek

In het verlengde van het bewustzijn over de effecten van AI is ethiek de laatste paar jaar een belangrijke pijler geworden in de discussie. Overheden en bedrijven ontwikkelen ethische codes en richtlijnen voor een verantwoorde inzet van AI en ook bij universitaire opleidingen worden vakken op het gebied van ethiek toegevoegd aan voorheen zuiver technische curricula.[192] De laatste tijd is er bij zowel technische als sociale studies ook aandacht voor de bredere

190 Keymolen et al. 2020.
191 Vetzo et al. 2018.
192 Bij de AI-opleiding in Delft (TU Delft) is ethiek één van de onderzoeksdomeinen, in Eindhoven (TUe) vormt ethiek een verplicht vak en in Amsterdam (UvA) is de cursus 'Fairness, Accountability, Confidentiality & Transparency' onderdeel van het lesprogramma.

relatie tussen AI en de samenleving.[193] Naast de benoeming van technische AI-hoogleraren zijn er nu ook leerstoelen voor hoogleraren die zich richten op de sociale en maatschappelijke kant van AI. Als faculteitshoogleraar AI, Data & Democratie houdt Claes de Vreese (UvA) zich bezig met de manier waarop AI democratische processen beïnvloedt en Tibor Bosse (RU) richt zich als hoogleraar Communicatiewetenschap & Kunstmatige Intelligentie op de interactie tussen mensen en intelligente mediatechnologie.

De structurelere bestudering van de implicaties van AI haakt aan bij een zekere 'ethiekmoeheid'. Nu hebben veel zaken die in de praktijk onder die noemer worden verricht, weinig van doen met ethiek, maar er groeit ontevredenheid over de talrijke ethische codes en richtlijnen die voor AI worden ontwikkeld. Ze blijken vaak ver af te staan van de complexiteit van de praktijk en onvoldoende houvast te bieden om misstanden en ongewenste ontwikkelingen te voorkomen. Er lijken structurelere waarborgen nodig om ervoor te zorgen dat AI in lijn is met onze waarden, en dat duwt de aandacht voorbij de specifieke toepassingen naar de bredere dynamiek van maatschappelijke inbedding.

Aandacht voor de inbedding van AI in de samenleving

We bevinden ons nu op het punt dat er brede aandacht is voor AI als multi-inzetbare technologie en het economisch potentieel daarvan. Ook is duidelijk geworden dat het gebruik van AI een transformerende uitwerking zal hebben op bestaande praktijken en mogelijk leidt tot onwenselijke situaties. Die aandacht heeft zich tot recent gericht op de korte termijn en vaak op specifieke waarden. Inmiddels heeft de discussie zich verbreed, naar de effecten die AI binnen verschillende domeinen heeft en naar de impact op een breder palet aan waarden. Waar de AI-discussie aanvankelijk vooral over privacy, transparantie en menselijke controle ging, is er nu ook aandacht voor de uitwerking van AI op andere waarden, zoals duurzaamheid (zie tekstbox 2.3).

Tekstbox 2.3 – AI en duurzaamheid

Door de intrede van AI in de samenleving neemt de aandacht toe voor de maatschappelijke gevolgen ervan. Vraagstukken rondom privacy, gelijke behandeling, autonomie en veiligheid worden steeds meer bediscussieerd. Een onderwerp dat nog maar beperkt op de maatschappelijke

193 De AI-opleidingen in Utrecht (UU) en Amsterdam (VU) kennen respectievelijk de vakken 'Philosophy of AI' en 'AI & Society', en aan de sociale faculteit van Maastricht University wordt sinds kort de bachelor 'Digital Society' aangeboden, waarin de impact van digitale technologieën zoals AI op de samenleving centraal staat.

en politieke radar staat, maar waar steeds meer onderzoek naar wordt gedaan, is het effect van AI op duurzaamheid.

Verschillende optimisten verwachten dat AI veel kan bijdragen aan duurzaamheid. De jaarlijkse VN-conferentie AI for Good richt zich onder andere op ecologische doelstellingen in relatie tot AI. Inmiddels zijn er ook al veel concrete projecten waarbij AI bijdraagt aan duurzaamheid. Welbekend zijn de initiatieven voor het efficiënter gebruiken van energie en het beter voorspellen van wind- en zonne-energie. Maar ook wordt AI gebruikt voor *smart farming*. De Amsterdamse start-up Connecterra gebruikt algoritmes van Google voor de veeteelt.[194] En er zijn interessante initiatieven rondom natuurconservatie. eBird gebruikt bijvoorbeeld machine learning-algoritmes voor vogelaars en de data die die algortimes genereren, worden gebruikt in campagnes om vogels te beschermen. Global Fishing Watch gebruikt AI om de visserij te monitoren. Ten slotte mag het EU-initiatief Destination Earth (DestinE) niet onvermeld blijven, waarin ook de inzet van AI is voorzien.[195]

Daartegenover is er steeds meer bewijs dat AI ook een negatieve invloed op duurzaamheid heeft. De CO_2-voetafdruk van de mondiale computinginfrastructuur is al groter dan die van de luchtvaart op het hoogtepunt van die industrie. Het draaien van een enkel algoritme voor *natural language processing* genereert evenveel uitstoot als 125 heen- en weervluchten van New York naar Beijing.[196] Ook wordt AI ingezet om nog meer fossiele energie te produceren en niet-duurzame consumptie te stimuleren.

Dauverge betoogt dat er voor elk positief voorbeeld meer voorbeelden te geven zijn waarbij AI niet duurzaam is. Hij wijst hier op de achterliggende politieke economie en machtsstructuren rondom de technologie. Zolang die gekenmerkt worden door een commerciële logica gericht op exploitatie, zal AI geen positieve bijdrage leveren aan duurzaamheid.[197] Veranderingen aan de machtsstructuren, partijen en doelen waarvoor de technologie wordt ingezet, zijn nodig om het potentieel van AI voor duurzaamheid te kunnen ontsluiten.

194 Dauverge 2020.
195 Europese Commissie z.d. (a).
196 Crawford 2021.
197 Dauverge 2020.

Ook de adviesaanvraag van de regering aan de WRR geeft duidelijk weer hoe de vraag naar de impact van AI alle beleidsterreinen betreft en potentieel alle waarden raakt die daarbinnen centraal staan. De aandacht voor de effecten van AI in specifieke contexten heeft zich als het ware geaggregeerd tot de vraag naar de impact van AI op de samenleving als geheel.

Nadat AI als revolutionaire technologie op de maatschappelijke agenda is komen te staan, zijn de effecten en dan met name de risico's van AI gaandeweg een belangrijker aandachtspunt gaan vormen in het publieke debat. Nu AI op steeds meer plaatsen in de praktijk wordt, en zal worden, toegepast, wordt ook de vraag naar haar effecten structureler: hoe ervoor te zorgen dat we AI zo ontwikkelen dat we daarbij datgene dat we als samenleving belangrijk vinden, onze publieke waarden, bewaken? Die vraagt dwingt ons om de blik voorbij de directe uitwerkingen van AI te richten op de langere termijn. Niet alleen onze technische systemen moeten robuust worden ingericht om publieke waarden te beschermen, dat geldt ook voor de samenleving zelf.

Op dit punt in de ontwikkeling van AI staan we voor precies die opgave – om te bepalen wat nodig is om deze technologie structureel in te bedden in onze samenleving. Voordat we uitgebreid ingaan op de verschillende aspecten daarvan, is het belangrijk om eerst stil te staan bij de rol die het lab tijdens dat proces zal blijven spelen.

Kernpunten – AI als maatschappelijk fenomeen

- Met de overstap van lab naar samenleving is er een maatschappelijke dynamiek rondom AI op gang gekomen.
- In de eerste plaats kenmerkt die dynamiek zich door aandacht voor AI als revolutionaire technologie. Het zijn met name de economische kansen die in het begin centraal staan.
- Naarmate er in de praktijk meer ervaring wordt opgedaan met AI, wordt ook duidelijker welke negatieve consequenties de inzet ervan kan hebben. Naast aandacht voor de kansen groeit de aandacht voor de risico's en er ontstaat een publieke en politieke discussie over AI.
- De AI-discussie is aanvankelijk gericht op bepaalde waarden, zoals privacy, non-discriminatie en transparantie, en de toepassing in specifieke contexten. De brede toepasbaarheid van AI beweegt de discussie uiteindelijk echter naar de impact van AI op de samenleving als geheel en alle publieke waarden daarbinnen.

2.4 De toekomst van het lab

We hebben gezien dat AI het lab uit gaat en zich op allerlei manieren in de samenleving mengt – in de vorm van een veelheid aan toepassingen en via het maatschappelijk debat daarover. Wat betekenen deze ontwikkelingen voor de toekomst van het lab? Dat AI definitief haar intrede heeft gedaan in de samenleving, betekent niet dat het lab in belang of dynamiek afneemt. Dat zou een misvatting zijn. Met de transitie van lab naar samenleving is AI nog niet 'af' en het is dus ook niet zo dat we ons vanaf nu vooral met toepassingen bezig moeten houden.[198] Dat AI nu op zo'n brede manier vanuit het lab haar weg naar de samenleving vindt, betekent dat zich nu in feite *extra* vraagstukken voordoen rondom de inbedding ervan in de samenleving. We lichten die opgaven in het volgende hoofdstuk toe.

Ondanks de grote activiteit op het gebied van concrete toepassingen, blijft ontwikkeling in het lab echter cruciaal om ten minste twee redenen. Dat de doorbraken van de laatste jaren veel innovatie mogelijk maken, laat ten eerste onverlet dat AI ook nog aanzienlijke beperkingen kent. De huidige methoden bieden op veel vragen nog geen antwoord en er wordt bijvoorbeeld nu al gewezen op de grenzen van de mogelijkheden van in het bijzonder deep learning. Of dat tot een derde AI-winter gaat leiden valt niet te zeggen, maar zeker is wel dat AI nog een lange weg te gaan heeft en dat significante voortgang meer fundamenteel onderzoek vergt. De tweede reden waarom het lab van belang blijft, heeft te maken met het specifieke karakter van AI. Het is een type technologie waarbij het lab bij de toepassing betrokken blijft. Strikt genomen verlaat AI het lab dus nooit echt.

De noodzaak van fundamenteel onderzoek

Diverse deskundigen merken op dat toegang tot meer en betere data de sleutel vormt tot het overwinnen van allerlei huidige beperkingen van machine learning. Er zijn bovendien interessante ontwikkelingen binnen het veld van machine learning, zoals het gebruik van *generative adversarial networks* (GANs). Hierbij worden verschillende algoritmes gebruikt om elkaar te verbeteren. Het ene algoritme genereert iets nieuws, zoals een door het algoritme zelfgemaakte afbeelding van een vogel, waarna het andere algoritme aangeeft

198 Die suggestie doet de computerwetenschapper Kai-Fu Lee wel (Lee, 2018: 143). Uiteraard erkent hij dat AI nog veel verder door kan ontwikkelen, maar hij meent dat de ontdekkingen van de laatste jaren zo groot zijn geweest, dat het onwaarschijnlijk is dat er binnenkort opnieuw doorbraken van die orde uit het lab voortkomen. Hij stelt bijvoorbeeld dat er in meer dan een decennium sinds het belangrijke paper van Geoffrey Hinton niets in machine learning is ontwikkeld dat even revolutionair was. In zijn ogen moeten we daarom onze aandacht meer op toepassingen gaan richten dan op fundamenteel onderzoek. Dat is niet de boodschap van dit rapport.

of het de afbeelding als een vogel herkent of niet. Zo niet, dan gaat het eerste algoritme door totdat het tweede algoritme 'overtuigd' is.

Ray Kurzweil meent dat met dit soort methoden van simulatie allerlei problemen van gebrekkige data zijn op te lossen. In plaats van zelfrijdende auto's te laten leren op de weg met alle gevaren die dat meebrengt, kunnen ze in gesimuleerde werelden miljoenen kilometers maken zonder iemand in gevaar te brengen.[199] Op eenzelfde manier kunnen bij defensie robots worden getraind binnen een simulatie, waardoor ze al verder ontwikkeld zijn wanneer ze in de fysieke wereld moeten opereren. Een andere veelbelovende aanpak is *federated learning*. Bij deze aanpak worden data niet opgenomen in een centrale server om een machine learning-algoritme te trainen, maar worden algoritmes verbeterd door hun parameters aan te passen met die van andere datasets zonder die data te combineren. Een uitkomst voor privacygevoelige data, zoals die van ziekenhuizen.

Ondanks dergelijke ontwikkelingen merken wetenschappers op dat innovatie noodzakelijk blijft, onder meer omdat er inherente begrenzingen aan machine learning lijken te kleven. Een voorbeeld daarvan doet zich voor op het gebied van computer vision. Vooruitgang op dat gebied is noodzakelijk om op termijn autonome voertuigen mogelijk te maken of in te kunnen zetten voor veiligheidsdoeleinden. Huidige algoritmes kunnen echter vrij gemakkelijk voor de gek gehouden worden. In een aantal experimenten bleek dat het plaatsen van hele kleine afbeeldingen van verkeersborden – te klein voor het menselijk oog – de algoritmes van een zelfrijdende auto deden handelen alsof het normale verkeersborden waren. Algoritmes zoeken volgens de huidige technieken namelijk naar patronen en de heel precieze eigenschappen van een dergelijk bord leverden voor het algoritme een voldoende mate van zekerheid op dat het officiële verkeersborden betrof. Dat de afbeeldingen bijzonder klein waren en in deze context waarschijnlijk niet klopten, was niet iets dat tot de algoritmes 'doordrong'.

Zo'n aanval – een *adversarial attack* – op het gedrag van algoritmes kan in het geval van zelfrijdende auto's desastreuze gevolgen hebben. Een recente studie liet zien dat het aanpassen van een enkele pixel een AI-algoritme kan misleiden. Bij gebruik in het leger kan dit eveneens grote gevolgen hebben. Een andere studie liet namelijk zien hoe een classificator van beelden om de tuin geleid kon worden door een machinegeweer als een helikopter te identificeren.[200] Er zijn ook andere toepassingen van dit soort aanvallen. Google gebruikt een

199 Uit een interview met Ray Kurzweil (Ford 2018: 230).
200 Tonin 2019: 6.

algoritme om video's te classificeren, onder meer om intellectueel eigendom te beschermen. Onderzoekers van de University of Washington toonden aan dat dit algoritme om de tuin te leiden valt door willekeurige beelden voor een fractie van een seconde in te voegen bij video's.[201] In de VS dook er een voorbeeld op van een politieagent die muziek afspeelde terwijl hij werd gefilmd, waarschijnlijk in de hoop dat de algoritmes van YouTube de video vanwege de schending van intellectueel eigendom zouden blokkeren.[202]

Oppervlakkig en inefficiënt

Tal van andere tekortkomingen van machine learning tonen eveneens dat er nog veel werk in het lab nodig is, zo betogen ook AI-pioniers zelf. Yoshua Bengio stelt dat diepe neurale netwerken statistische regelmatigheden aan de oppervlakte van datasets leren, maar geen abstracte concepten van een hogere orde. Daardoor ontbreekt een vorm van begrip die nodig is voor bepaalde communicatie en taken. Geoffrey Hinton en Demis Hassabis, de oprichter van DeepMind, hebben beiden gezegd dat *general artificial intelligence* nog niet in de buurt is van een werkbare realiteit.[203]

Pionier Hinton sprak zich kritisch uit over huidige methoden en benoemde verschillende tekortkomingen.[204] Een van die tekortkomingen is inefficiëntie. In tegenstelling tot oudere benaderingen lijkt machine learning dichter bij het menselijk leervermogen te zitten. Net als bij mensen worden beelden herkend door patronen te herkennen in plaats van door vaste regels te volgen. Tegelijkertijd zijn er grote verschillen en kunnen mensen dit proces veel efficienter uitvoeren. Een klein kind heeft na het zien van een paar appels al een goed vermogen om toekomstige appels te herkennen. De machine learning-algoritmes moeten getraind worden met duizenden beelden van appels om ze goed te kunnen herkennen. Ook al neemt de hoeveelheid beschikbare data wereldwijd toe, dat gebeurt niet in alle domeinen en vaak is er een tekort aan goede data. Bovendien kan het voor bepaalde toepassingen te riskant zijn om eerst veel fouten te maken voordat het algoritme goed getraind is.

Gezond verstand

Een probleem dat daarmee samenhangt, is het ontbreken van gezond verstand (*common sense*). Dat blijkt uit het bovengenoemde voorbeeld van de kleine afbeeldingen van verkeersborden. Huidige algoritmes zijn ingesteld op alle mogelijke afbeeldingen en kunnen daarmee niet goed onderscheiden wat in een

201 Agrawal et al. 2018: 200.
202 Thomas, 9 februari 2021.
203 Marcus en Davis 2019: 62.
204 Dickson, 2 maart 2020.

bepaalde context waarschijnlijk is en wat niet. Ook al herkennen ze patronen, ze zijn niet goed in staat om daar betekenis aan toe te kennen. Een voorbeeld hiervan zijn CAPTCHA's: Completely Automated Public Turing tests to tell Computers and Humans Apart. Dit zijn de testen die mensen online weleens moeten uitvoeren om te bewijzen dat ze geen bots zijn, door bijvoorbeeld alle foto's aan te klikken waarop bomen te zien zijn. Dat berust op ons vermogen om gezond verstand te gebruiken. Zelf als er maar een fragment van een boom zichtbaar is, kan een mens vaak gemakkelijk afleiden of iets een boom is uit de locatie of objecten eromheen, zoals een struik.

Voor een algoritme zijn die beperkte datapunten voor herkenning al snel een probleem. De CAPTCHA toont de beperkingen van de huidige machine learning in situaties waar gezond verstand vereist is. ML-algoritmes kunnen niet leren van de collectieve kennis die eerder elders door andere programma's is opgedaan. Daardoor hebben ze moeite met vragen waar mensen meteen antwoord op kunnen geven, zoals "Wie is langer, prins William of zijn babyzoon prins George?" en "Als je een speld in een wortel steekt, komt er dan een gat in de speld of in de wortel?"[205]

Dit hangt samen met het feit dat mensen vaak leunen op allerlei impliciete kennis. Als wij iets zeggen, noemen we niet alle relevante informatie, maar gaan we ervan uit dat anderen die kennis uit de context afleiden. Als iemand tegen een taxichauffeur zegt "breng mij zo snel mogelijk naar het vliegveld", dan weet de chauffeur dat dat niet betekent "ten koste van de levens van anderen of het volgen van de verkeersregels". Dat is impliciete achtergrondkennis die een algoritme niet heeft. Anders gezegd, de taal is ondergespecificeerd en geen enkel feit is een eiland.[206] Stuart Russell geeft het voorbeeld van de vooruitgang op het gebied van natuurkundige kennis. Door telescoopdata te analyseren kan een algoritme nieuwe kennis ontwikkelen. Maar vooruitgang in de natuurkunde volgt niet alleen uit het bestuderen van meer data. Het opzetten van hypothesen en het bedenken welke factoren in het universum van alle data wel of niet meegenomen moeten worden, volgt uit eerdere kennis van de natuurkunde, kennis die niet in data besloten ligt.[207]

Ondoorzichtigheid

Een andere tekortkoming is de ondoorzichtigheid van huidige ML-algoritmes. Het is vaak bijzonder moeilijk om te traceren hoe een algoritme tot een bepaalde uitkomst is gekomen. In veel gevallen is dat proces wel transparant te maken,

205 Marcus 2018: 12.
206 Marcus en Davis 2019: 136-139.
207 Russell 2019: 83.

maar het is daarmee nog niet uitlegbaar: inzage betekent nog geen inzicht. Een beslissing om iets te classificeren op basis van eigenschappen op het niveau van pixels is voor mensen bijvoorbeeld niet te volgen. Dat is in veel gevallen geen groot probleem, maar wanneer het gaat over bijvoorbeeld het identificeren van een veiligheidsrisico of een beslissing die veel impact heeft op iemands leven, zoals het verstrekken van een hypotheek of het stellen van een medische diagnose, is dat zeer problematisch. Voor dat soort toepassingen is uitlegbaarheid een vereiste.

Volgens sommige experts is de complexiteit van dit soort algoritmes geen onoverkomelijk probleem. Eerdergenoemde Hassabis stelt dat we momenteel de systemen aan het bouwen zijn en dat daarna een proces volgt van terugredeneren via *reverse engineering* om te begrijpen hoe die systemen werken. Hij meent dan ook dat binnen een decennium de meeste systemen geen '*black box*' meer zullen zijn.[208] Yann LeCun maakt de vergelijking dat we de verbrandingsmotor aan het uitvinden zijn en ons nu al zorgen maken over de remmen en veiligheidsgordels. Dat zijn volgens hem problemen die later spelen en opgelost zullen worden.[209]

Andere experts menen echter dat gebrekkige uitlegbaarheid inherent is aan deze techniek en om een andere benadering vraagt. Volgens Judea Pearl groeit menselijke kennis niet door een blind proces, maar door het maken en testen van modellen van de werkelijkheid. Huidige machine learning-benaderingen zijn beperkt omdat zij gericht zijn op correlatie, niet op causaliteit.[210] Hij maakt een analogie met het verschil tussen de Babylonische en de Griekse astronomie. De eerste kon vaak zeer precieze voorspellingen doen, beter dan de Griekse, maar was onnavolgbaar – een *black box*: de mechanismen achter de voorspellingen werden niet begrepen. De Griekse benadering was wel gericht op begrip van die mechanismen en die nadruk op causaliteit bleek centraal in de latere opkomst van de wetenschap. Pearl stelt dat de huidige niet-modelmatige benadering van machine learning om die reden tekortschiet.[211]

(Oude) nieuwe benaderingen

Om aan al deze tekortkomingen tegemoet te treden wordt er gewerkt aan andere benaderingen binnen AI. Een aantal daarvan bouwt voort op de *good old-fashioned AI*, de symbolische AI waar het veld mee begon. Dit soort regelgebaseerde systemen wordt bijvoorbeeld gebruikt in gevallen waar niet veel

208 Uit een interview met Demis Hassabis (Ford 2018: 178).
209 Uit een interview met Yann LeCun (Ford 2018: 136).
210 Uit een interview met Judea Pearl (Ford 2018: 363).
211 Pearl 2019: 18.

data beschikbaar zijn. Siemens gebruikt een dergelijk systeem om processen in de gasturbines van fabrieken aan te sturen. Zonder vooraf geprogrammeerde regels zou machine learning de gasturbines een eeuw moeten laten draaien om even effectief te worden.[212] Ook op gebieden waar het de privacy van mensen betreft, is het moeilijk om machine learning op grote hoeveelheden data los te laten met uitkomsten die ondoorzichtig zijn. Hier zouden eveneens *top-down* logische systemen een oplossing kunnen zijn. Een verwante denkrichting is het gebruik van regelgebaseerde systemen *naast* machine learning om uitkomsten ervan te voorspellen en zo te kunnen deduceren welke regels gevolgd worden en uitkomsten dus uitlegbaar te maken. Mensen als Yann LeCun en Nick Bostrom menen dat de toekomst ligt in het toevoegen van structuur en modellering in bestaande machine learning-technieken.[213]

In een variant op deze hybride benadering wordt gewerkt aan het coderen van gezond verstand in algoritmes. DARPA heeft bijvoorbeeld een Machine Common Sense-programma. Daarbij worden modellen gemaakt die net als bij menselijke cognitie een onderscheid maken tussen verschillende categorieën, zoals objecten, plaatsen en actoren. Verwant zijn benaderingen waarin algoritmes in plaats van alles vanaf het begin blanco te moeten leren al bepaalde principes meekrijgen. Een klein kind leert immers ook via 'inductieve biases' die al vroeg in het brein zitten. Op jonge leeftijd leert het kind de basisfysica over het bestaan van objecten, hoe zij door de ruimte bewegen en dat ze bijvoorbeeld niet door elkaar heen kunnen gaan. Dergelijke principes sturen het leerproces waardoor het veel sneller kan gaan en een kind dus niet duizenden voorbeelden moet zien om iets te herkennen. Een benadering die hiervan uitgaat, is de *graph network*, waarin objecten als bollen en relaties als lijnen worden gerepresenteerd.[214] Geoffrey Hinton werkt aan 'gedachtenvectoren' om de betekenis van taal beter te vatten[215] en ook het Project Mosaic van het Allen Institute for AI werkt aan het programmeren van gezond verstand in computers.

Weer andere benaderingen zitten zeer dicht tegen de neurowetenschap aan. Net als neurale netwerken geïnspireerd zijn op de werking van het brein, zijn er initiatieven om *neuromorphic chips* te maken. De chips van computers worden dan gemodelleerd op het functioneren van zenuwcellen. Het Human Brain Project van de Europese Unie heeft de ambitie om met computers een brein na te bouwen.[216] Dat is ook het doel van het Amerikaanse BRAIN-initiatief.[217]

---

212 Wilson et al., 14 januari 2019.
213 Ford 2019: 78, 108, 126.
214 Waldrop 2019.
215 Marcus en Davis 2019: 128.
216 Marsh, 10 januari 2019.
217 Domingos 2015: 118.

Margaret Boden onderscheidt naast de twee bekende benaderingen van symbolische AI en artificiële neurale netwerken nog drie andere benaderingen: *evolutionary programming, cellular automata* en *dynamical systems*.[218] Ook Pedro Domingos geeft in zijn zoektocht naar het ultieme algoritme aan dat er naast de symbolische en neurale benaderingen nog drie andere benaderingen zijn die aan die zoektocht zullen bijdragen. *Genetic programming* is een benadering die wordt gebruikt bij het ontwerp van elektronica en het optimaliseren van fabrieken.[219] Daarnaast zijn er Bayesiaanse methoden, zoals *naive Bayes classifiers* en *hidden Markov models,* die gebruikt worden bij onder andere spamfilters, spraakherkenningssystemen en het opschonen van datareeksen.[220] Een laatste benadering bestaat uit systemen die op analogie zijn gebaseerd, zoals het *nearest neighbor*-algoritme of *support vector machines.* Deze benadering wordt gebruikt bij wetenschappelijke modellen van het zonnestelsel of van atomen en bij het maken van muziek in de stijl van een bepaalde componist.[221]

Ondanks dat *machine learning* de afgelopen jaren een enorme vlucht heeft genomen en in rap tempo op allerlei domeinen wordt toegepast, zijn er dus beperkingen aan deze benadering en wordt er onderzoek gedaan naar alternatieven. Er zijn afruilen tussen de krachten en zwakten van verschillende benaderingen, waardoor ze voor verschillende toepassingen beter geschikt zijn. Het is goed mogelijk dat het in de toekomst zal draaien om de vraag op welk gebied welke benadering het beste past, in plaats van om een enkele benadering die in alle gevallen superieur is. Veel experts voorzien dan ook een toekomst van hybride benaderingen en beweren dat menselijke intelligentie net zo werkt. Het onbewust herkennen van bekende patronen is bijvoorbeeld te begrijpen aan de hand van neurale netwerken, terwijl in onbekende situaties bewust geredeneerd wordt, wat meer overeenkomt met symbolische AI. Al met al kan dus geconcludeerd worden dat AI allerminst uitontwikkeld is en dat fundamenteel onderzoek van groot belang zal blijven.

Het lab hoort bij AI

De tweede reden waarom het lab in de toekomst een belangrijke rol zal blijven spelen, heeft zoals gezegd te maken met het karakter van AI zelf en breder met het karakter van digitale technologie. Terwijl traditioneel iets in een lab wordt uitgevonden, het vervolgens in een fabriek tot eindproduct wordt gemaakt en dan wordt verkocht aan een klant, kennen digitale producten een andere dynamiek. De ontwikkelaar ervan blijft namelijk bij de toepassing betrokken. Neem

218 Boden 2018: 5-6.
219 Domingos 2015: 133.
220 Domingos 2015: 151-155.
221 Domingos 2015: 199.

het verschil tussen traditionele televisie en streamingdiensten als Netflix. Een televisieproducent zendt een programma uit, waarna het de latere feedback van kijkers kan gebruiken om een nieuw product te verbeteren. Bij Netflix blijven we op het platform van het bedrijf, dat *realtime* leert van onze data en daarmee het platform direct kan aanpassen. Anders gezegd, in plaats van eindproducten zijn digitale producten altijd *halffabricaten*. Ze zijn niet af en worden in het gebruik continu verbeterd en aangepast, of het nu gaat om een streamingplatform, een slimme thermostaat of een gezondheidsapp.

Dat sluit aan bij de werkwijze van technologiebedrijven. Zogenoemde *'lean start-ups'* werken snel toe naar de lancering van een '*minimal viable product*' (MVP). Een MVP werkt vaak nog niet goed en heeft nog allerlei problemen (*'bugs'*), maar heeft wel een minimum aan levensvatbaarheid. In de praktijk kunnen dit soort systemen vervolgens verder leren en verbeteren. Dat betekent ook dat 'het lab' in zekere zin aanwezig blijft in de toepassing en een grote rol speelt om die toepassing in de praktijk vorm te geven. Niet voor niets ontstaan er in Nederland allerlei samenwerkingsverbanden zoals de genoemde AI-labs, waarbij wetenschappers (het 'lab') gekoppeld worden aan bedrijven of overheidsdiensten. AI heeft dus niet alleen de overstap gemaakt van lab naar samenleving, we zouden ook kunnen zeggen dat het lab zelf de samenleving ingaat bij AI. Een andere manier om dit voor te stellen is dat de samenleving het lab in is getrokken en een '*living lab*' wordt. Voor de ontwikkeling van nieuwe diensten doet Facebook bijvoorbeeld continu allerlei experimenten met gebruikers van het platform. Dat leidde overigens tot grote controverse over experimenten waarbij Facebook de emoties van mensen probeerde te beïnvloeden.[222]

Deze specifieke dynamiek rondom de ontwikkeling van digitale producten brengt allerlei vraagstukken met zich mee die we in latere hoofdstukken zullen tegenkomen. Ze betekent bijvoorbeeld dat het in een later stadium moeilijker is om de oorspronkelijke problemen van het MVP te verhelpen, een fenomeen dat '*technical debt*' heet.[223] Ze kan ook allerlei risico's met zich meebrengen. Doordat een product niet eerst als eindproduct wordt getest, kunnen gebruikers aan iets worden blootgesteld dat onwenselijke of zelfs schadelijke gevolgen voor ze heeft. Als we daar eenmaal achter komen, is het kwaad vaak al geschied.

Het karakter van AI als halffabricaat levert ook uitdagingen op voor toezichthouders. Overheidsdiensten zoals de Rijksdienst voor het Wegverkeer (RDW) controleren auto's voordat ze de Nederlandse weg op mogen. Dat kan goed met eindproducten, maar werkt minder goed voor halffabricaten die veranderen

222 Zie onder andere NOS, 29 juni 2016, en Rutkin, 25 juni 2014.
223 Marcus en Davis 2019: 188.

als ze zich al op de weg bevinden, zoals een Tesla die een software-update krijgt. Er wordt in dit kader dan ook gekeken naar de ontwikkeling van een soort 'continue APK'. Ook in de zorg geldt dat apparaten op functionaliteit en veiligheid worden getest voor goedkeuring, waarbij het uitgangspunt is dat de werking daarna in de kern niet meer verandert. Maar in het geval van AI past het systeem zich voortdurend aan en kunnen van een afstand veranderingen worden doorgevoerd. Deze 'lab-dynamiek' vraagt dus ook om een dynamischer benadering van de beoordeling van dit soort systemen. In deel 2 van dit rapport komen we op dit vraagstuk terug. Voor nu is het belangrijk om in te zien hoe het lab in het geval van AI nauw betrokken blijft in de praktijk en daar niet zondermeer van kan worden onderscheiden.

In de toekomst van AI zal het lab dus een grote rol blijven spelen. De grenzen van huidige benaderingen vragen om verder fundamenteel onderzoek en de aard van AI zorgt ervoor dat het lab altijd verbonden blijft met de toepassing in de praktijk. Voor de vooruitgang van de technologie en de bevordering van de inpassing ervan in de samenleving is het dus zaak om de functie van het lab niet uit het oog te verliezen, het lab te betrekken in de praktijk en uit te rusten met voldoende middelen en talent.

Kernpunten – De toekomst van het lab

- Ook al gaat AI van het lab de samenleving in, dat betekent niet dat het lab minder relevant wordt. Dat heeft twee belangrijke redenen:
- Ten eerste kleven er allerlei tekortkomingen aan huidige AI-methodieken. Die zijn oppervlakkig, inefficiënt, missen gezond verstand en zijn ondoorzichtig. Fundamenteel onderzoek blijft dus van groot belang om deze tekortkomingen te adresseren.
- Ten tweede geldt voor AI, net als voor bredere digitale technologie, dat het onderzoek van het lab verbonden blijft met de praktijk. Het lab gaat als het ware zelf de samenleving in.

3. AI als systeemtechnologie

Nu we weten wat AI is en gezien hebben hoe de technologie de laatste jaren uit het lab de samenleving in is gegaan, zoomen we in op dit inbeddingsproces. Wat is ervoor nodig om AI in te passen in onze samenleving? Voor de beantwoording van die vraag ontwikkelen we in dit hoofdstuk een kader, waarin we AI beschouwen als een speciaal type technologie, namelijk een *systeemtechnologie*, met enkele historische precedenten. Door AI op die manier te begrijpen kunnen we lessen trekken uit de geschiedenis van andere technologieën van dat type. Vanuit dat perspectief kunnen we nadenken over wat ons te doen staat met AI en hoe we kunnen omgaan met de vele vraagstukken waar zij ons voor stelt. Dat betekent overigens niet dat de geschiedenis zich herhaalt of dat de ontwikkeling van technologie iets deterministisch heeft. We ontwikkelen geen rigide kader, maar identificeren brede patronen die perspectief bieden op het heden, zonder daarbij de verschillen met het verleden uit het oog te verliezen. Met deze benadering kunnen we voorbij de actualiteit kijken en daarmee ten dele ook voorbij de waan van de dag.

Verschillende prominente figuren hebben in hun typering van AI parallellen getrokken met andere technologieën. Volgens Andrew Ng, de onderzoeker die we in het vorige hoofdstuk tegenkwamen, is de impact van AI "te vergelijken met die van elektriciteit een eeuw geleden."[224] Google's CEO Sundar Pichai en zijn voorganger Eric Schmidt vergeleken AI eveneens met elektriciteit. Pichai durfde het zelfs aan om de vergelijking met vuur te maken.[225]

Ook in de op beleid georiënteerde wereld komen we dergelijke parallellen regelmatig tegen. In een paper over de strategische implicaties van AI schrijft Michael Horowitz bijvoorbeeld dat zij geen geïsoleerde technologie is, maar lijkt op breed toepasbare technologieën als elektriciteit en de verbrandingsmotor.[226] Die breedte van mogelijke toepassingen beaamt ook het European Political Strategy Center van de Europese Commissie: "Het is moeilijk om een segment van de samenleving voor te stellen dat in de komende jaren niet door AI getransformeerd zal worden."[227]

224 Lynch, 4 mei 2017.
225 Goode, 19 januari 2018; Morozov 2013: 1.
226 Horowitz et al. 2018.
227 European Political Strategy Centre 2018.

De Nederlandse overheid onderscheidt al langer technologieën met een brede toepassing en spreekt in dit verband van 'sleuteltechnologieën'.[228] Bij de lancering van de Nederlandse AI-strategie stelde een van de participanten echter dat AI "meer is dan een sleuteltechnologie" en beter begrepen kan worden als een "basistechnologie, vergelijkbaar met elektriciteit en de verbrandingsmotor".[229]

In veel uitingen en documenten wordt dus gehint op overeenkomsten met eerdere technologieën, maar die vergelijking wordt vrijwel nergens nader uitgewerkt. Die opdracht nemen we daarom in dit hoofdstuk ter hand. Daartoe bekijken we welke implicaties de vergelijking van AI met technologieën als elektriciteit heeft. We bespreken de literatuur over verschillende typen technologieën, in het bijzonder het idee van *general purpose technologies,* en munten de term 'systeemtechnologie'. Uit de historische ontwikkeling van systeemtechnologieën leiden we een aantal algemene patronen af over de maatschappelijke inbedding ervan. We onderscheiden vijf specifieke opgaven die deel uitmaken van dit inbeddingsproces. In deel 2 van dit rapport passen we deze opgaven van maatschappelijke inbedding stuk voor stuk toe op AI.

3.1 De typering van technologieën

Er wordt al lang onderzoek gedaan naar hoe verschillende typen technologieën brede invloed hebben op de economie en samenleving. Een vroege notie daarvan is te vinden in het idee van Kondratieff-golven en vooral bij Joseph Schumpeter, die op dat idee voortborduurde. Schumpeter constateerde dat perioden van hogere economische groei zich afwisselden met perioden van lagere groei en schreef dat toe aan het effect van nieuwe technologieën. Periodiek verhogen sets van verschillende nieuwe technologieën de groei, waarvan het effect na verloop van tijd weer terugloopt. Die dynamiek was volgens hem inherent aan de kapitalistische markt: "het essentiële punt om te begrijpen wanneer we het over het kapitalisme hebben, is dat we met een evolutionair proces te maken hebben", (…) "industriële mutatie", (…) dat de economische structuur continu van binnenuit revolutionair verandert, zonder te stoppen de oude structuur vernietigt en een nieuwe maakt."[230] Dit is zijn beroemde idee van 'creatieve destructie'.

In zijn toespraak bij het in ontvangst nemen van de Nobelprijs voor de Economie in 1971, sprak de econoom Simon Kuznets van zogenoemde 'epochale innovaties' die een tijdperk van grote economische ontwikkeling aandrijven.

---

228 Het document *Aanpak Sleuteltechnologieën* van de Rijksoverheid (Kamerstukken II 2018/19, 33009, nr. 70) schaart AI onder de categorie van 'hightech'. De term 'sleuteltechnologie' is ontwikkeld door de nationale High-Level Group.

229 Tijdens de presentatie van het Strategisch Actieplan voor AI op 8 oktober 2019.

230 Citaat van Joseph Schumpeter in Juma 2016: 17.

De innovatiewetenschappers Carlota Perez en Chris Freeman hebben het over een vergelijkbaar fenomeen, dat zij 'nieuwe technologiesystemen' en 'technologische revoluties' noemen.[231] Deze vormen een krachtig en zeer zichtbaar cluster van nieuwe en dynamische technologieën, producten en industrieën die in de hele economie tot grootscheepse verandering leiden en op lange termijn tot economische groei. Sinds de Industriële Revolutie onderscheidt Carlota Perez vijf van dat soort clusters, waaronder de tijd van stoom en spoorwegen, van staal en elektriciteit, van olie, auto's en massaproductie, en de tijd van informatie- en telecommunicatie. Elke technologische revolutie brengt volgens haar een eigen 'techno-economisch paradigma' met zich mee, een manier van denken en doen, waardoor de technologie zich verweeft met de hele samenleving.[232] Een relevante aanvulling op dit idee komt van Alessandro Nuvolari, die benadrukt dat grote effecten niet zozeer komen van individuele technologieën, maar eerder van blokken van radicale innovatie die tezamen de revolutie uitmaken.[233] Innovatie bestaat volgens sommige onderzoekers dan ook niet in de ontwikkeling van iets nieuws en groots, maar gaat over het maken van combinaties tussen zaken die al langer bestaan.[234]

General purpose technologies

Vanuit deze ideeën is AI te typeren op basis van de brede, transformerende impact ervan op de samenleving. Als we naar de technologie zelf kijken is het concept van *'general purpose technologies'* (GPTs) interessant. Dit zijn technologieën die niet slechts een beperkte toepassing hebben, zoals grasmaaiers, broodroosters of een microscoop, maar die een generiek karakter hebben en in talloze vormen kunnen worden toegepast voor uiteenlopende doeleinden. Daardoor kan dit soort technologieën door de hele economie en samenleving heen grote invloed uitoefenen. Timothy F. Bresnahan en Manuel Trajtenberg introduceerden het concept in 1992 in een artikel.[235] Daarin benoemen zij drie criteria voor GPTs. Ten eerste hebben GPTs een hoge mate van alomtegenwoordigheid (*pervasiveness*) en verspreiden ze zich over veel sectoren, productieprocessen en producten. Ten tweede hebben ze een groot potentieel voor technische verbetering (*technical improvements*), waardoor kosten blijven dalen en de efficiëntie van de technologie stijgt. Ten slotte leiden GPTs tot allerlei complementaire innovaties (*innovational complementarities*), waardoor de bredere productiviteit in de economie toeneemt.

231 Freeman en Louçã 2001: 144.
232 Perez 2003: 8-11. Ook de Belgische econoom Luc Soete is actief in dit veld.
233 Nuvolari 2016.
234 Brynjolfsson en McAfee 2014: 78.
235 Bresnahan en Trajtenberg 1995.

Inmiddels heeft zich een uitgebreide literatuur opgebouwd over het concept van GPTs. Dat betekent echter niet dat overal dezelfde definitie wordt gehanteerd[236] of dat de term op dezelfde manier wordt toegepast. Sommige auteurs gaan uit van slechts enkele historische GPTs, terwijl anderen een lange lijst ontwikkelen door ver terug te gaan in de tijd, waarbij ze ook de domestificatie van dieren en de bewerking van brons als vroege voorbeelden beschouwen. Eén auteur komt tot 28 technologieën die in de literatuur als GPT zijn bestempeld.[237]

Een ander punt van discussie in de literatuur is dat er ook technologieën zijn geweest met een enorme maatschappelijke invloed, maar die desondanks niet bijzonder generiek zijn. Denk aan de drukpers of het stoomschip, twee technologieën die maar voor een beperkt aantal taken ingezet kunnen worden, terwijl ze de samenleving zeer radicaal veranderd hebben. De meest gehanteerde voorbeelden van GPTs waar de meeste overeenstemming over bestaat, zijn de stoommachine, elektriciteit, de verbrandingsmotor en ICT.[238]

Ondanks de voornoemde kanttekeningen zijn er de laatste jaren interessante studies uitgevoerd die AI expliciet vanuit dit concept van GPTs benaderen. We noemen er een paar. Naar aanleiding van een door het Amerikaanse National Bureau of Economic Research (NBER) georganiseerde conferentie in 2017 verscheen in 2019 de bundel *The Economics of Artificial Intelligence*. Het eerste deel van de bundel is getiteld '*AI as a* GPT' en bevat bijdragen van gerenommeerde onderzoekers van technologie, evenals van vooraanstaande economen. De bundel bevat interessante analyses waar we hier gebruik van zullen maken, maar is in lijn met de oorspronkelijke conferentie primair gericht op de macro-economische effecten van AI, wat niet de focus van voorliggend rapport is.

Een andere interessante studie is het proefschrift van Jade Leung aan Oxford University getiteld: *Who will govern artificial intelligence? Learning from the history of strategic politics in emerging technologies*. In deze studie onderzoekt Leung AI naast ruimtevaarttechnologie, biotechnologie en cryptografie als vier voorbeelden van wat ze 'strategische GPTs' noemt. Vanuit dat perspectief legt zij de nadruk op de relatie tussen overheden en nieuwe technologieën en in het bijzonder op de rol van de wereld van defensie. Leung onderscheidt daarbij de overheid, bedrijven en onderzoekers als actoren. Ze laat zien dat de doelen,

---

236 Gavin Wright (2000) definieert GPTs bijvoorbeeld als "diepe nieuwe ideeën of technieken die het potentieel hebben voor significante invloed op veel sectoren van de economie".

237 Zie hiertoe het Working Paper *Artificiële intelligentie als een general purpose technology – Strategische belangen van verantwoorde inzet in historisch perspectief* (Bakker en Korsten 2021) dat Freedomlab in opdracht van de WRR heeft uitgevoerd.

238 Voor een kritische analyse van het verschillend gebruik van de term 'GPT', zie Field 2008.

instrumenten en beperkingen van deze actoren verschillen, in bepaalde fasen kunnen botsen en in andere fasen kunnen convergeren.

Naast deze bronnen is er ook recent onderzoek waarin AI wordt getypeerd vanuit een breed, historisch perspectief zonder daarbij expliciet de term GPT te gebruiken. In polemiek met het befaamde boek *The Second Machine Age* van Andrew MacAfee en Erik Brynjolfsson heeft de eerdergenoemde Carlota Perez een negendelige serie artikelen geschreven getiteld *'Second Machine Age or Fifth Technological Revolution?'*. Daarin onderzoekt zij op welke manier we hedendaagse digitale technologie – waaronder AI – met eerdere technologieën kunnen vergelijken.[239]

Deze studies, en vooral het perspectief dat daarin wordt ontwikkeld, zijn relevant voor het theoretisch kader vanwaaruit we AI in dit rapport benaderen. Daarnaast putten we uit concretere studies van de effecten van specifieke technologieën. Sarah A. Seo schreef over de beroemdste toepassing van de verbrandingsmotor, de auto. Zij laat daarbij zien hoe dit symbool van vrijheid tegelijkertijd heeft geleid tot een enorme vergroting van de macht van de staat en specifiek van de politie in het private leven van burgers.[240] In een breed overzicht van een reeks technologieën van tractoren en margarine tot elektriciteit en GMO onderzocht Calestous Juma de dynamiek van maatschappelijk verzet tegen nieuwe technologieën.[241] Naast de studie van GPTs, putten we in dit onderzoek dus ook uit analogieën met recente technologieën als GMO en nanotechnologie, die interessante parallellen met AI vertonen.[242]

AI als GPT

Vanuit het voorgaande ligt vervolgens de vraag op tafel of AI inderdaad als een GPT te beschouwen is. Een bevestigend antwoord daarop lijkt vrij gemakkelijk te geven, al zitten we pas in een vroeg stadium van de wijdverbreide impact van AI. Als we kijken naar de drie eigenschappen van GPTs die Bresnahan en Trajtenberg formuleerden, zijn die overtuigend van toepassing op AI.

De eerste is *alomtegenwoordigheid*. Ook al komt de verspreiding van AI in de economie en bredere samenleving pas de laatste jaren echt op gang, de technologie wordt nu al in heel diverse sectoren en producten gebruikt. Eerder in dit deel van het rapport (paragraaf 2.2) hebben we uiteenlopende voorbeelden

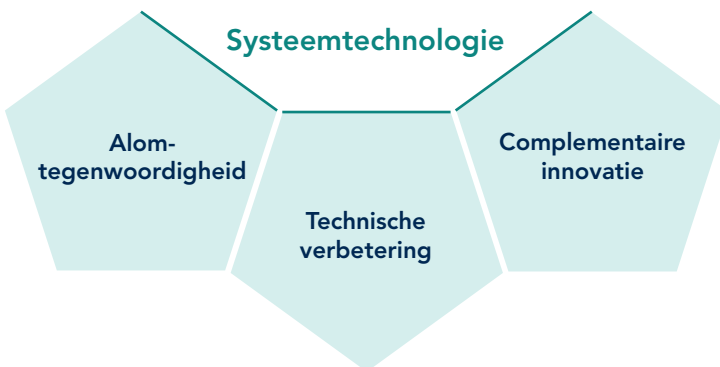
239 Perez 2017-2020.
240 Seo 2019.
241 Juma 2016.
242 In het rapport *Betekenis van Nanotechnologieën voor de Gezondheid* beschrijft de Gezondheidsraad deze technologie bijvoorbeeld vanuit het idee van een *'enabling technology'*, wat op een interessante manier aansluit bij onze bespreking van GPTs.

besproken van AI in onder andere de industrie, landbouw, publieke sector, entertainment, financiële instellingen en medische praktijk. Het is nu al duidelijk dat AI door deze brede inzetbaarheid op weg is naar alomtegenwoordigheid in de hele samenleving en economie.

De tweede eigenschap is het *potentieel voor technische verbetering*, met lagere kosten en een hogere efficiëntie tot gevolg. Ook dat is duidelijk aan te tonen voor AI. In hoofdstuk 1 beschreven we hoe rekenkracht volgens de Wet van Moore elke twee jaar verdubbelt en hoe die verdubbeling verdere verbetering van AI-technieken mogelijk maakt. We zagen ook hoe wetenschappelijk onderzoek tot een proces van nieuwe en verbeterde technieken heeft geleid. Als gevolg daarvan werden er bij de toepassing van AI de laatste jaren allerlei mijlpalen bereikt. Uit de bespreking van de toekomst van het lab blijkt bovendien dat er nieuwe veelbelovende technieken worden ontwikkeld die de prestaties en efficiëntie van AI verder kunnen verbeteren.

Ten slotte leidt een GPT tot *complementaire innovatie* die de algemene productiviteit verhoogt. Er zijn in het geval van AI al verschillende voorbeelden die wijzen op hogere productiviteit, maar om dit overtuigend te kunnen aantonen bevinden we ons simpelweg nog in een te vroeg stadium. In hun prognoses gaan gezaghebbende onderzoeksinstituten en consultancybureaus als Accenture, PwC, McKinsey en Deloitte desalniettemin uit van grote productiviteitsverhogingen over het komende decennium. In figuur 3.1 zetten we de drie kenmerkende eigenschappen van systeemtechnologieën op een rijtje.[243]

Figuur 3.1 Drie eigenschappen van systeemtechnologieën



243 Voor een uitgebreide analyse van AI als GPT, zie Bakker en Korsten 2021.

AI als systeemtechnologie

We concluderen dat AI de drie eigenschappen van een *general purpose technology* heeft. Het concept van GPTs en de rijke literatuur waarin AI als zodanig wordt onderzocht, biedt waardevolle aanknopingspunten om te begrijpen met wat voor technologie we te maken hebben. Toch zullen we hier niet de term 'GPT' gebruiken, maar kiezen we ervoor om AI te typeren als *systeemtechnologie*. In onze analyse leggen we namelijk andere accenten dan in de literatuur over GPTs wordt gedaan.

Ten eerste kent de literatuur van GPTs, van vroege bronnen bij Kondratieff tot aan de genoemde recente studie van het Amerikaanse NBER, een sterke focus op de macro-economische effecten van deze technologieën. Veel onderzoekers proberen de effecten van technologieën te kwantificeren. Dat roept discussies op over de vraag of en hoe aangetoond kan worden dat GPTs tot een langdurig hogere economische groei leiden. Gezien de enorme hoeveelheid variabelen waarmee rekening gehouden moet worden, leidt dat tot complexe modellen. Ons accent ligt elders. Wij richten de aandacht niet primair op de kwantitatieve, maar op de kwalitatieve veranderingen die een systeemtechnologie als AI teweegbrengt.

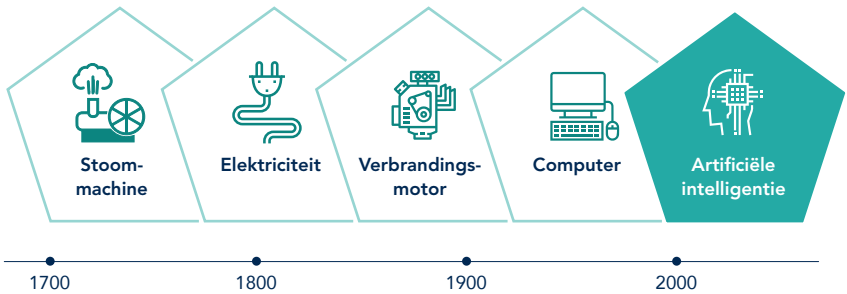
Ten tweede is er in de literatuur over GPTs prominent aandacht voor historische indelingen. Zoals we gezien hebben, is er veel discussie over het aantal GPTs die we historisch kunnen ontwaren. De ene onderzoeker telt er tientallen, Perez onderscheidt vijf clusters, auteurs als Chandler spreken van drie Industriële Revoluties[244], Schwab identificeert er vier, en Brynjolfsson en McAfee hebben het over twee 'machinetijdperken'. Verschillende van deze auteurs werken bovendien met sterk schematische weergaven met precieze begin- en einddata van bepaalde technologieën. In dit rapport onthouden wij ons van dergelijke indelingen en daarmee onderscheiden we ons van de GPT-literatuur. Juist omdat we ons vooral op de kwalitatieve en niet zozeer op de kwantitatieve effecten richten, behoeven we ons niet te committeren aan een strikte indeling met vaste begin- en eindpunten. Veeleer willen we bredere patronen in beeld brengen. Daarvoor richten we ons primair op een beperkt aantal eerdere systeemtechnologieën – te weten de stoommachine, elektriciteit, de verbrandingsmotor en de computer – en putten wij pragmatisch uit de historie voor relevante parallellen.

Bovendien legt de term 'GPT' de nadruk op de eigenschap dat een technologie voor veel doelen gebruikt kan worden. Met 'systeemtechnologie' willen wij dat accent verleggen naar enerzijds het *systeemkarakter* van bepaalde technologieën en anderzijds de blik verbreden naar de *systemische effecten* ervan op

244 Freeman en Louçã 2001: 145.

de samenleving. 'Systeem' verstaan wij dus tweeledig. Allereerst bestaat de technologie zelf uit een systeem met verschillende componenten. Elektriciteit functioneert in samenhang met generatoren, kabels en batterijen. Ook AI is onderdeel van een breder technisch systeem van data en hardware. Daarnaast doelen wij met systeemtechnologie op het effect dat een dergelijke technologie heeft op allerlei systemen en processen in de samenleving. Dat effect brengt een complex proces van aanpassing, oefening en onderhandeling met zich mee. Anders geformuleerd, deze term brengt het proces van inbedding in de samenleving in beeld en de vooral kwalitatieve effecten die daarmee gepaard gaan.

Figuur 3.2 AI als een nieuwe systeemtechnologie



Overeenkomsten en verschillen met eerdere systeemtechnologieën

AI is dus een systeemtechnologie en als zodanig te vergelijken met eerdere technologieën van dat type, zoals de stoommachine, elektriciteit en de verbrandingsmotor (figuur 3.2). Wel kunnen we AI daarbinnen nog nader karakteriseren. AI heeft namelijk bepaalde eigenschappen waardoor ze op sommige punten beter met de ene technologie te vergelijken is dan met de andere. Terwijl een verbrandingsmotor en een stoommachine tastbaar zijn, heeft AI net als elektriciteit een zeker immaterieel karakter. Eigenstandig bestaat het niet, maar slechts als onderdeel van een product of dienst. In deze zin zijn objecten als een broodrooster, lamp of radio *met* elektriciteit, te vergelijken met thermostaten, horloges en machines *met* AI.

Een andere onderscheiding die AI dichter bij elektriciteit dan de verbrandingsmotor plaatst, is het verschil tussen technologieën die '*technology-radical*' en '*use-radical*' zijn. De eerste categorie wordt primair door technische en wetenschappelijke ontwikkelingen gedreven. De nieuwsgierigheid van onderzoekers drijft ontwikkeling voort, zonder dat duidelijk is hoe en waarvoor de technologie gebruikt gaat worden. Bij de tweede categorie zijn de toepassingen vanaf het begin duidelijk en spelen commerciële factoren al vroeg een rol. De ontwikkeling is hier doelgericht. Dit geldt voor de verbrandingsmotor. AI is net

als elektriciteit lange tijd door onderzoekers voortgedreven zonder dat zij in die fase een idee hadden van de lucratieve toepassingen die wij nu kennen.

Een andere onderscheiding is te maken tussen systeemtechnologieën waar de overheid van meet af aan een vanzelfsprekende rol heeft en systeemtechnologieën waarvoor dit niet geldt. Sommige technologieën worden bijvoorbeeld specifiek voor defensiedoeleinden ontwikkeld en blijven van die sector afhankelijk voor verdere toepassingen en daarmee opdrachten tot ontwikkeling. Andere technologieën zijn meer '*civilian-first*', omdat zij vooral economische vooruitzichten brengen. Het vermogen van overheden om de ontwikkeling van een technologie te sturen is groter in de eerste categorie. Daartoe behoort onder andere de ruimtevaart, terwijl de biotechnologie een voorbeeld is van de tweede categorie. Beide zijn voorbeelden van wat Jade Leung strategische GPTs noemt.[245]

Voor AI geldt dat in een vroege fase vooral het Amerikaanse militaire instituut DARPA een belangrijke financier was. Desalniettemin bleef het onderzoek in die tijd fundamenteel en maakten militaire toepassingen maar een beperkt deel van het toepassingsbereik van de technologie uit. AI lijkt hierin dus meer op biotechnologie dan op de ruimtevaart. Ze verschilt echter van biotechnologie doordat bij de laatste de ontwikkelaars voor een zeer groot deel verbonden zijn aan grote (academische) laboratoria, terwijl innovatie op het gebied van AI op meer decentrale wijze verloopt. Dat heeft weer consequenties voor het vermogen van onderzoekers om onderling universele standaarden op te stellen.

Naast deze technische overeenkomsten en verschillen met andere systeemtechnologieën is het ook belangrijk om te kijken naar variatie in de maatschappelijke context en tijdgeest waarin een technologie geplaatst is. Neem de rol van de overheid. De stoommachine werd ontwikkeld in een laisser-faire klimaat in Engeland, waardoor de overheid maar een zeer beperkte rol speelde. De verbrandingsmotor en de auto daarentegen kwamen op in tijden van Keynesiaans overheidsbeleid met betrekking tot de economie. Terwijl overheden tegenwoordig via standaarden en wetgeving veel invloed op de economie uitoefenen, kwam AI op in een tijd waarin relatief veel weerstand bestond tegen een sterk sturende rol van de overheid in de economie. Dat is goed om te bedenken wanneer we willen leren van historische patronen bij andere systeemtechnologieën.

Een ander verschil in de maatschappelijke context van AI en eerdere technologieën betreft de mobilisatie van verschillende partijen in de samenleving. Deze hangt samen met de historische ontwikkeling richting meer welvaart en

245 Leung 2019.

democratisering, waardoor grotere groepen individuen in staat werden gesteld hun stem te laten horen in de publieke ruimte. Terwijl in het verleden ondernemers en overheden veel gemakkelijker hun stempel op de samenleving konden drukken, hebben in de huidige tijd het maatschappelijk middenveld, onderzoekers, individuen en de media een veel sterkere positie. Dat betekent ook dat zij eerder dan bij vroegere systeemtechnologieën gemobiliseerd zijn rondom AI en zich zullen mengen in de maatschappelijke inbedding van deze technologie.

Dit hangt samen met wat Trajtenberg de '*democratization of expectations*' noemt: ten tijde van de Industriële Revolutie was het vermogen van fabrieksarbeiders om voor zichzelf op te komen beperkt omdat de meesten van hen bezig waren te voorzien in hun eerste levensbehoeften. We komen hier in paragraaf 3.5 op terug, bij het voorbeeld van de Luddieten. In onze tijd participeren veel meer mensen in het openbare leven en zijn de belangen van werknemers beter vertegenwoordigd. Bovendien zullen mensen nu enerzijds minder geneigd zijn om de kosten van technologische verandering te dragen terwijl ze anderzijds hogere verwachtingen hebben ten aanzien van hun aandeel in de opbrengsten daarvan.[246]

De wereld is niet alleen meer gedemocratiseerd, maar ook meer geglobaliseerd dan in het verleden. Dat betekent dat vraagstukken rondom AI van het begin af aan ook meer een mondiaal karakter hebben. Daarbij gaat het zowel om de reikwijdte van markten, waardoor toepassingen van AI tot ver over de landsgrenzen impact kunnen hebben, als om de aanwezigheid van allerhande internationale kaders, zoals handelsverdragen, mensenrechten en technische standaarden, die het internationale verkeer in banen leiden. Interessant genoeg is de opkomst van eerdere systeemtechnologieën een belangrijke impuls geweest voor het ontstaan van allerlei nieuwe internationale organisaties die standaarden ontwikkelen[247], en die nu ook een rol spelen bij AI. Te denken valt onder meer aan organisaties op het terrein van de telecommunicatie en het internet, standaardisatieorganisaties als de ISO en internationale organisaties van ingenieurs als de IEEE. Ook in het verleden vond de ontwikkeling en inbedding van een technologie al deels plaats over landsgrenzen heen, maar de mate waarin dat gebeurt neemt door verdere globalisering met de tijd toe.

Een laatste punt van verschil met de context waarin eerdere systeemtechnologieën opkwamen, betreft de hogere graad van organisatie en communicatie onder wetenschappers. Terwijl die bij de stoommachine amper als groep georganiseerd waren, spelen academische organisaties, gedragscodes en standaarden een grote rol bij hedendaagse technologieën als AI.

246 Trajtenberg 2018: 178.
247 Kaiser en Schot 2014.

Het techno-economisch paradigma van AI

Tot slot kijken we naar Carlota Perez' idee van het techno-economische paradigma.[248] Dat concept houdt in dat een grote technologische verandering niet alleen tot andere producten en diensten leidt, maar ook bepaalde manieren van denken, werken en principes van organisatie met zich meebrengt. Neem de productie in fabrieken die opkwam met de Industriële Revolutie en de 'vernetwerkte' productie die mogelijk was door elektriciteit. Of denk aan de verbrandingsmotor, die niet alleen auto's bracht maar ook de lopende band. Fordisme, Taylorisme en *just-in-time*-productie zijn organisatieprincipes die daarvan zijn afgeleid. AI is nog volop in beweging en we kunnen het techno-economisch paradigma ervan dan ook nog niet definitief karakteriseren. Wel kunnen we vast een aantal contouren onderscheiden, die voortbouwen op eerdere vormen van digitalisering. Daar voegen we nieuwe elementen aan toe.

We ontwaren hier drie contouren van het techno-economisch paradigma van AI. De eerste betreft veranderingen in de aard van objecten en producten. Zoals we al bespraken aan het slot van het voorgaande hoofdstuk, is in het digitale domein niet zozeer sprake van eindproducten, maar eerder van halffabricaten: digitale producten zijn nooit af. In tegenstelling tot traditionele producten en diensten die op een gegeven moment de fabriek verlaten en verkocht worden, worden zaken in de digitale wereld continu herschreven en aangepast. Middels updates veranderen digitaal verbonden objecten zoals computers, maar ook auto's, camera's en medische hulpmiddelen continu. In Kevin Kelly's woorden blijft alles in een continue 'staat van wording'.[249] Of zoals Luciano Floridi stelt, worden *'things'* vervangen door *'-ings'*, verwijzend naar werkwoorden als *interact-ing, process-ing, network-ing, do-ing, be-ing.*[250]

Verbonden hiermee is dat fysieke objecten die een digitaal aspect krijgen, daarmee hun losstaande karakter verliezen. Adam Greenfield spreekt in deze zin van *porositeit* als overkoepelend karakter van onze hedendaagse technologieën. De grenzen tussen objecten, tussen gebruiker en platform, maar ook de muren van huizen, worden als het ware poreus doordat ze onderling zijn verbonden en in elkaar overlopen. Een veelheid aan actoren is zo betrokken en aanwezig in al die producten. Deze veranderingen in de aard van fysieke objecten brengt allerlei vraagstukken mee over veiligheid, privacy en verantwoordelijkheid.

Een tweede element van het technische paradigma van AI is – paradoxaal genoeg – dat terwijl objecten verbonden en individuen transparanter worden, juist veel van de technologie onzichtbaar wordt. Op een bijeenkomst in Davos in

248 Perez 2003.
249 Kelly 2017: 9-27.
250 Floridi 2014: 183.

2015 voorspelde Eric Schmidt dat het internet zou verdwijnen. Hij bedoelde niet dat het ten onder zou gaan, maar verwees naar een idee afkomstig uit een invloedrijk artikel van Mark Weiser uit 1991 getiteld *The Computer for the 21st Century*.[251] In dit artikel introduceerde Weiser het idee van '*ubiquitous computing*', een alomtegenwoordige architectuur van digitale technologie, en stelde: "De meest prominente technologieën zijn die technologieën die verdwijnen. Ze weven zichzelf in het weefsel van het dagelijks leven totdat zij er niet meer van te onderscheiden zijn." Hierdoor kunnen "computers naar de achtergrond verdwijnen".[252]

Luciano Floridi maakt hetzelfde punt met een metafoor. Volgens hem leven wij nu op de *piano nobile*, de centrale bovenverdieping van een Renaissancewoning die naar buiten toe zichtbaar is. Daaronder zijn echter allemaal bedienden, in ons geval digitale bedienden, aan het werk in de dienstkamers.[253] Interessant aan deze ruimtelijke metafoor is dat deze de aandacht vestigt op een verticale structuur. Verschillende lagen zijn bovenop elkaar geplaatst die niet allemaal zichtbaar zijn. Benjamin Bratton ziet in een verticale structuur de kern van digitale technologie.[254] Hiervoor leefden wij volgens hem in een horizontale wereld waarin mensen, objecten en landen naast elkaar op de wereldkaart stonden. Digitalisering brengt daar een verticale structuur in aan, met lagen van internetadressen, clouddiensten en datacenters die op een veelal ongemerkte manier door al die zaken heen lopen. In de wereld van technologie wordt vaak gesproken van een *'stack'* (stapel), een geheel dat uit verschillende gestapelde lagen van hardware, software, netwerk en applicaties bestaat. Het ontstaan van die goeddeels onzichtbare gelaagdheid roept vragen op ten aanzien van machtsrelaties en afhankelijkheden.[255] Jose van Dijck wijst via een andere metafoor op de verticale aard van digitale technologie. Zij spreekt van de boomstructuur van platformisering en richt daarmee de aandacht op machtsconcentratie door bijvoorbeeld verticale integratie.[256]

Een laatste element van het technische paradigma van AI dat wij hier noemen, bouwt voort op het technologieconcept van Floridi. Hij stelt dat het idee van technologie als een instrument problematisch is, omdat het suggereert dat een mens een instrument bedient en daarmee invloed uitoefent op een buitenwereld. Dat maskeert dat veel van onze technologie helemaal niet op een natuurlijke buitenwereld ingrijpt, maar op andere technologieën. Onze aandacht

---

251 Zuboff 2019: 200.
252 Weiser 1991.
253 Floridi 2014: 37.
254 Bratton 2016.
255 Het AI Now Institute doet ook uitgebreid studie naar de onzichtbare lagen van AI: van menselijke trainers van algoritmes in andere landen tot aan de materiële benodigdheden die leiden tot toeleveringsketens van allerlei grondstoffen. Zie bijvoorbeeld Joler en Crawford 2018.
256 Van Dijck 2020: 1-19.

moet juist uitgaan naar die 'intertechnologische' dynamiek. Floridi noemt technologieën die op technologieën ingrijpen 'tweede-ordetechnologieën'. Dat is bijvoorbeeld de rem die ingrijpt op de banden van de auto. Daarbij staat aan het begin wel een mens, die in dit geval de rem indrukt.

Wat AI echter mogelijk maakt, zijn 'derde-ordetechnologieën'. Technologieën die andere technologieën laten ingrijpen op weer andere technologieën zonder dat daar een mens aan te pas komt. De zelfrijdende auto zou bijvoorbeeld ook de beslissing kunnen nemen om te remmen. Overal waar AI-systemen autonoom tot beslissingen kunnen komen, kan de structuur van een derde-ordetechnologie ontstaan. Zo gaat het bijvoorbeeld al met bepaalde verkeersboetes. Een camera maakt een foto van een auto, stelt vast dat er sprake is van een snelheidsovertreding en zorgt ervoor dat een boete verstuurd wordt naar het adres van de eigenaar van de auto. Deze autonomie die technologie door algoritmes krijgt en die uitdrukkelijk onderdeel is van de definitie van AI die we in dit rapport hanteren, roept onder andere vragen op over menselijke controle, verantwoordelijkheid en het toerekenen van handelingen met rechtsgevolgen.[257]

Kernpunten – AI als systeemtechnologie

- Er is een brede literatuur die vernieuwende technologieën karakteriseert als '*epochal innovations*', '*technological revolutions*' en '*general purpose technologies*'. Een *general purpose technology* (GPT) onderscheidt zich door alomtegenwoordigheid, een groot potentieel voor technische verbeteringen en complementaire innovaties. AI voldoet aan alle drie criteria.
- We beschouwen AI in dit rapport als *systeemtechnologie*. In tegenstelling tot de literatuur over GPTs hanteren we geen rigide indeling en leggen we de nadruk op kwalitatieve eigenschappen en op het effect daarvan op de samenleving.
- Als systeemtechnologie is AI te vergelijken met onder andere de stoommachine, de verbrandingsmotor en elektriciteit. In bepaalde opzichten lijkt AI het meest op deze laatste technologie. Door de tijd heen zijn er veranderingen opgetreden in de maatschappelijke context waarbinnen dergelijke technologieën zich ontwikkelen.
- AI brengt een eigen techno-economisch paradigma met zich mee dat zich onderscheidt door de continue verandering van producten en diensten, een moeilijk zichtbare verticale ordening van apparatuur en software en de mogelijkheid van autonoom handelende technologie.

257 Hage 2017.

3.2 De inbedding van systeemtechnologieën

Met het begrip van AI als systeemtechnologie, kunnen we bekijken wat er nodig is voor de maatschappelijke inbedding ervan. Op basis van de geschiedenis van eerdere systeemtechnologieën identificeren we een aantal patronen in dat proces, waar we bij de omgang met AI lering uit kunnen trekken. In deze paragraaf bespreken we enkele algemene lessen, voordat we in de volgende paragraaf ingaan op vijf specifieke opgaven waar een samenleving zich met een nieuwe systeemtechnologie voor geplaatst ziet.

Een co-evolutie tussen samenleving en technologie

Om te beginnen betreft een proces van inbedding een *langdurige co-evolutie* van samenleving en technologie. Zo'n proces vraagt oefening, experiment en onderhandeling – en dat kost tijd. Dat plaatst het sterk gepolariseerde debat rondom AI meteen in perspectief. Zo zijn er techniekoptimisten die menen dat AI op korte termijn de maatschappij fundamenteel kan verrijken met zelfrijdende auto's, geavanceerde medische diagnoses en geautomatiseerde productie. Sceptici daarentegen beschouwen de technologie als een hype en wijzen erop dat concreet bewijs ontbreekt dat AI nu al groot effect sorteert en dat allerlei verwachtingen, zoals die van de introductie van de zelfrijdende auto, steeds verder naar de toekomst verschuiven.

In beide standpunten zit een kern van waarheid. Zoals de optimisten verwachten, zal AI veel nieuwe mogelijkheden creëren. Zij verkijken zich echter op het gemak waarmee die nieuwe mogelijkheden maatschappelijk ingebed kunnen raken. Daarvoor wijzen de sceptici terecht op problemen in de voorzienbare toekomst, hoewel die geen algehele scepsis over de technologie rechtvaardigen. Een systeemtechnologie vraagt om een proces van wederzijdse adaptatie van samenleving en technologie en dat vergt tijd, zelfs in onze tijd van snelle technologische ontwikkeling en globalisering. Technologieën kunnen misschien sneller de wereld over gaan, maar het inbedden ervan, het zorgen dat ze werken en dat mensen ze vertrouwen, doet een beroep op maatschappelijke processen die niet per se sneller gaan dan in het verleden. Een dergelijk proces gaat met horten en stoten en duurt decennia.

Onvoorspelbare ontwikkeling en effecten

Een hiermee samenhangende observatie is dat de introductie van een nieuwe systeemtechnologie in hoge mate een *onvoorspelbaar proces* is. Nieuwe technologieën worden vaak voor andere dingen gebruikt dan waarvoor ze initieel werden benut of bedoeld. Don Ihde spreekt in deze zin van 'multistabiliteit' en Wiebe Bijker van 'interpretatieve flexibiliteit'.[258] Auto's werden oorspronkelijk

258 Verbeek 2014: 31.

gebruikt voor sport en medische doeleinden, omdat de ijle lucht goed was voor de longen.[259] Thomas Edison had de grammofoon niet gemaakt voor vermaak, maar als een apparaat voor het zakenleven, een soort dictafoon die hij een *'talking machine'* noemde.[260]

Het niet goed kunnen voorzien hoe een technologie gebruikt gaat worden en de daarmee samenhangende onderschatting van de effecten ervan, noemt Shoshana Zuboff het '*horseless carriage syndrome*'.[261] Grote technologische revoluties brengen iets onvoorstelbaar nieuws en worden begrepen vanuit wat mensen al wel kennen. De auto werd initieel dan ook voorgesteld als een koets zonder paarden. Door hem als een meer efficiënte versie te zien van iets dat al bestond, onderschatten mensen destijds de gevaren van een auto en de invloed die deze uiteindelijk op de samenleving zou gaan hebben. Dit klinkt achteraf misschien als een naïeve voorstelling van zaken, maar als we nu spreken over 'zelfrijdende auto's' doen we misschien hetzelfde en bezien we een nieuwe technologie als een verbeterslag op iets bestaands, waardoor we de werkelijke invloed ervan niet in het vizier krijgen.

De introductie van de auto begin twintigste eeuw ging overigens gepaard met nog een andere misvatting. Het idee was namelijk dat auto's zouden helpen om de vervuiling van steden op te lossen.[262] Het vervoer met paarden zorgde voor grote hoeveelheden mest in de stad en daardoor voor stank en de verspreiding van ziekten. Daartegenover werd de auto gezien als een snellere en schonere manier van vervoer. Door mest uit de stad te verwijderen, maakte de auto de stad inderdaad schoner en leefbaarder. Wat mensen zich destijds echter onvoldoende realiseerden, was dat de auto zelf weer andere vormen van vervuiling en leefbaarheidsproblemen met zich mee zou brengen. Dit voorbeeld laat zien dat nieuwe technologieën *vaak onbedoelde bijeffecten* hebben. De introductie van waterleidingen en riolen in huizen beoogde ziekten tegen te gaan, maar verloste tegelijkertijd huisvrouwen van een van de meest zware taken in het huis: het naar binnen en buiten slepen van water.[263] Een onbedoeld bijeffect van elektrisch licht was dat het aantal sterfgevallen afnam, omdat er daarvoor veel ongelukken gebeurden met olielampen.[264]

---

259 Verbeek 2014: 71.
260 Gordon 2016: 186.
261 Zuboff 2019: 12. Zuboff beschrijft bovendien dat bedrijven soms bewust iets radicaals nieuws als iets ouds voorstellen zodat burgers het gaan gebruiken. Denk aan de surveillancetechnologie van Google Glass, die de vorm kreeg van een normale bril (Zuboff 2019: 156).
262 Bakker en Korsten 2021.
263 Gordon 2016: 123.
264 Gordon 2016: 237.

Daar komt bij dat technologische veranderingen ook tot gedragsveranderingen kunnen leiden die een tegengesteld effect hebben dan oorspronkelijk bedoeld. Terwijl spaarlampen ontwikkeld waren om minder energie te verbruiken, hebben ze uiteindelijk tot hogere energieconsumptie geleid, omdat ze gebruikt werden op plaatsen waar voorheen geen verlichting was, zoals in tuinen.[265] Edward Tenner spreekt in dit verband over het '*rebound effect*' van technologieën.[266] Terwijl nieuwe apparaten in het huis de taken van huisvrouwen minder zwaar maakten, ging dat ook gepaard met een verhoging van standaarden rondom bijvoorbeeld schone kleding, waardoor juist meer huishoudelijk werk werd gecreëerd.[267]

De onvoorspelbaarheid van systeemtechnologieën zit ook in de structurele veranderingen die ze teweegbrengen op de lange termijn en die niet te voorzien zijn. Neem het effect dat spoorwegen hadden op de ruimtelijke ordening van de stad, doordat werknemers niet meer op loopafstand van hun werk hoefden te wonen. Of denk aan het effect dat de auto had op de jeugdcultuur van de jaren zestig en de nieuwe uitgaansgelegenheden die hij mogelijk maakte, zoals de drive-inbioscoop, het drive-throughrestaurant, het motel en wegrestaurants.[268]

Al deze bronnen van onvoorspelbaarheid hebben consequenties voor het proces van inbedding van AI. We zullen er rekening mee moeten houden dat veel ontwikkelingen niet te voorzien zijn. Grote terughoudendheid is vereist bij zeer stellige toekomstscenario's of bij het maken van lineaire extrapolaties uit het verleden. Die houden onvoldoende rekening met de onverwachte effecten van een nieuwe technologie.

De impact op publieke waarden

Die les ligt ook ten grondslag aan onze keuze om de vraag naar de impact van AI op publieke waarden te beantwoorden vanuit structurele maatschappelijke opgaven. Op welke wijze publieke waarden door een systeemtechnologie geraakt zullen worden, valt namelijk onmogelijk bij voorbaat af te bakenen en is bovendien vaak niet eenduidig. Dat blijkt uit bovengenoemde voorbeelden van het effect van de auto op de leefbaarheid, van elektriciteit op de emancipatie en van de trein op het samenleven. Juist vanwege de brede toepasbaarheid van een systeemtechnologie is er geen uitsluitsel te geven over de publieke waarden die in het geding zijn – potentieel gaat het over het hele palet aan waarden. Zo zal het ook zijn met AI. AI zal invloed hebben op onder andere veiligheid en

265 Verbeek 2014: 22.
266 Tenner 1997.
267 Gordon 2016: 278.
268 Gordon 2016: 166.

gezondheid, autonomie en vrijheid, burgerrechten en de rechtsstaat, rechtvaardigheid en inclusie, maar op manieren die we nu nog niet kunnen voorspellen.

Desalniettemin worden er momenteel veel inventarisaties gemaakt van waarden, principes en rechten waar AI invloed op heeft. Op die manier worden actuele vraagstukken in kaart gebracht en dat is van belang om daarop te kunnen acteren. Tegelijkertijd is het ook nodig om te kijken voorbij dit moment en de manier waarop AI nu bepaalde waarden beïnvloedt, juist als we die waarden op de lange termijn willen borgen. Met onze benadering van maatschappelijke inbedding vullen we de huidige discussie aan door de aandacht te richten op het langdurige proces waarbij samenleving en technologie elkaar beïnvloeden en in theorie alle publieke waarden in het spel zijn.

Regels en succes zijn geen vijanden

Een laatste algemene observatie bij de inbedding van systeemtechnologieën is dat er *geen inherente spanning* is tussen publieke waarden en de bijbehorende regels enerzijds en het economische succes van de technologie anderzijds. Dat is een veelgehoorde mythe in de context van AI die in het volgende hoofdstuk specifieker aan de orde komt. Een blik op historische technologische revoluties leert dat de tegenstelling tussen normatieve kaders en innovaties niet terecht is. Uiteraard kunnen regels technologische ontwikkeling remmen, door iets expliciet te verbieden, zoals nucleaire technologie voor militaire doeleinden. Veel regels en normen helpen echter om een technologie betrouwbaarder te maken. Dat vergroot de bereidwilligheid van burgers en bedrijven om die technologie te gaan gebruiken en stimuleert zo juist de omarming ervan.

Neem opnieuw het gebruik van de auto. Daar is over de jaren een complex stelsel van regels en normen voor ontwikkeld in de vorm van keurmerken (APK), toezichthoudende instanties (Rijkswaterstaat, RDW), veiligheidseisen (gordels, airbags, aanwezigheid van reservebanden), hulpdiensten (ANWB), verplichte verzekeringen en natuurlijk de veelheid aan verkeersregels met het bijbehorende rijexamen. Al deze normatieve kaders verhinderen het gebruik van de auto niet, maar stimuleren het juist doordat ze gevaren verkleinen en zekerheid vergroten. Zonder APK, gordels, verzekering, airbags en verkeersregels was autorijden veel riskanter en waarschijnlijk minder populair geweest. Daar komt bij dat het proces van normering en regulering van de auto na decennia nog steeds voortduurt.

Hetzelfde patroon doet zich voor bij de spoorwegen. De eerste treinen waren gevaarlijk, oncomfortabel en vies. De houten stoelen waren ongemakkelijk, het rook er naar voedsel en tabak, en reizigers waren bij aankomst doortrokken van een vieze laag rook. R.L. Stevenson noemde in een essay de trein "de ark

van Noach op wielen".[269] Stap voor stap werd de treinrit door regels en normen echter veiliger en prettiger. Een andere interessante analogie is de opkomst van industrieel geproduceerd voedsel in de negentiende eeuw. Ook dat was aanvankelijk een gevaarlijk en ongezond Wild Westen. Zonder keurmerken, toezichthouders en wetgeving waren burgers vaak de dupe van dubieuze praktijken. Veel voedsel werd bijvoorbeeld aangelengd. Kalk en gips werd toegevoegd aan melk om de gele kleur witter te maken en vervuild water in de melk leidde tot de verspreiding van tuberculose en tyfus.[270]

AI zal een vergelijkbaar patroon volgen. De technologie vindt haar weg naar de samenleving maar de regels, normen en praktijken die AI daar inbedden, moeten voor een groot deel nog ontwikkeld worden. Ook hier is sprake van een Wild Westen met allerlei risico's, voor individuen en de samenleving als geheel. Dat is geen argument tegen het gebruik van de technologie, maar een oproep om te werken aan een lang proces waarbij AI op verantwoorde wijze in de samenleving gebruikt kan worden.

Kernpunten – De inbedding van systeemtechnologieën

- Systeemtechnologieën brengen een langdurig proces van co-evolutie tussen samenleving en technologie met zich mee en kunnen de samenleving daarmee ingrijpend veranderen.
- De ontwikkeling van systeemtechnologieën is vaak onvoorspelbaar en de effecten ervan zijn van tevoren niet te overzien.
- Welke publieke waarden door een systeemtechnologie beïnvloed worden, valt niet af te bakenen. Door het generieke karakter zijn potentieel alle publieke waarden in het geding.
- Er is geen inherente spanning tussen normen en wetten enerzijds en de verdere ontwikkeling en toepassing van nieuwe technologieën anderzijds.

Naast deze algemene patronen bij de inbedding van systeemtechnologieën, onderscheiden we vijf specifieke opgaven die de hoekstenen van dit proces vormen.

269 Gordon 2016: 141.
270 Gordon 2016: 220.

3.3 Opgave 1: Demystificatie

Over grasmaaiers en broodroosters bestaan geen mythen. Hun doel en werking is vrij duidelijk en ze laten daarom weinig aan de verbeelding over. Dat is anders bij systeemtechnologieën. Juist door hun generieke karakter hebben die iets ongrijpbaars, waardoor er gemakkelijk mythen ontstaan die weinig met de werkelijkheid van doen hebben. Enerzijds ontstaan er allerlei onrealistisch hoge verwachtingen over waar de nieuwe technologie allemaal toe in staat is en hoe ze een bijna magisch antwoord vormt op allerlei maatschappelijke vraagstukken. Anderzijds ontwikkelen zich ook veel extreme angstbeelden en doemscenario's over de effecten van de nieuwe technologie. De eerste categorie mythen kan gemakkelijk leiden tot teleurstelling en de tweede tot afkeer. Beide leiden er in ieder geval toe dat de verkeerde vragen en kwesties worden geadresseerd ten aanzien van de nieuwe technologie. Bij de maatschappelijke inbedding van systeemtechnologieën is het daarom van belang om te werken aan een realistisch beeld van wat de technologie wel en niet kan en wat voor effecten ze heeft. De vraag die demystificatie adresseert, is daarom: waar hebben we het over? (figuur 3.3)

Figuur 3.3 Opgave 1: Demystificatie

Verschillende maatschappelijke actoren zijn bij deze opgave betrokken. Omdat het mede de publieke beeldvorming betreft, is het bredere publiek bij uitstek bij deze opgave betrokken. Bedrijven die de nieuwe technologie ontwikkelen, vervullen door hun marketing vaak een rol bij het creëren van te hoge verwachtingen. Concurrerende bedrijven met andere technologieën of in meer traditionele sectoren kunnen een rol spelen bij het creëren van angst voor de technologie. Ook partijen in het maatschappelijk middenveld kunnen de mythen benadrukken, juist omdat zij gericht zijn op mogelijke risico's. De overheid ten slotte heeft vaak een belang bij het gebruik van nieuwe technologieën,

en kan daardoor bijdragen aan overhaast enthousiasme. Tegelijkertijd kan zij de beeldvorming ook negatief beïnvloeden door met haar handelen bepaalde associaties bij een nieuwe technologie te versterken.

Hooggespannen verwachtingen

Hoe zijn de patronen van demystificatie terug te zien in de geschiedenis van systeemtechnologieën? Kijkend naar optimisme is duidelijk dat sinds de Industriële Revolutie nieuwe technologieën geassocieerd worden met vooruitgang en beschaving. Elektriciteit bijvoorbeeld werd een "*defining element of a great cilivization*" en inspireerde tot allerlei utopische boeken.[271] Door elektrisch licht kreeg Berlijn de bijnaam 'stad van licht'. Elektriciteit werd niet alleen geassocieerd met de eerdergenoemde emancipatie maar ook met schoonheid, flexibiliteit en de algemene verbetering van levensomstandigheden. Het was een voorbeeld van het breder fenomeen van sciëntisme – het idee dat wetenschappelijke vooruitgang leidt tot maatschappelijke vooruitgang – en geloof in de maakbaarheid en zelfs perfectionering van de samenleving. Op een manier die doet denken aan de verwachtingen rondom AI, adverteerde General Electric (GE) in 1917 al haar elektronische apparaten als "elektrische bedienden" die werkten "zonder te klagen".[272]

Een ander voorbeeld van hoge verwachtingen dat relevant is voor de hedendaagse discussies over digitalisering, is het idee dat deze nieuwe technologieën vrede brengen. De negentiende-eeuwse ingenieur Michel Chevalier beschreef de spoorwegen als "het belangrijkste middel voor vrede in Europa en geluk voor de mensheid".[273] Van de telegraaf werd verwacht dat deze "harmonie tussen mensen en naties" zou faciliteren en door de mensheid te verenigen barrières van "vooroordelen en gebruik" zou doorbreken.[274] Ook Henry Ford, pionier van de auto, keek in de jaren twintig op die manier naar de moderne industrie. Het is interessant om hem uitgebreid te citeren:

> "Machinerie bereikt in de wereld wat de mens niet lukte met preken, propaganda of het geschreven woord. Het vliegtuig en de draadloze media kennen geen grenzen. Ze bewegen over de gestippelde lijnen op de kaart zonder aandacht of moeite. Zij brengen de wereld bij elkaar op een manier die geen enkel ander systeem kan. De film met zijn universele taal, het vliegtuig met zijn snelheid, de draadloze media met zijn aanstaande internationale programma – deze zullen

271 Bakker en Korsten 2021: 16.
272 Gordon 2016: 120.
273 Van der Vleuten et al. 2017: 27.
274 Gordon 2016: 178.

> binnenkort volledig begrip in de wereld brengen. We zouden dus een Verenigde Staten van de Wereld kunnen voorzien. Die zal uiteindelijk zeker komen!"[275]

Er zijn altijd specifieke kanalen geweest waarlangs utopische beelden zich verspreiden. Een eerste is natuurlijk de sciencefictionliteratuur. Welbekend is de roman van schrijver William Gibson, *Neuromancer*. Hierin geeft hij een lyrische beschouwing van een nieuwe wereld die hij als eerste met de term 'cyberspace' aanduidt.[276] Een ander kanaal zijn grote publieke wedstrijden. Innovatieve ondernemers gaan in de geschiedenis allerlei wedstrijden aan met oudere technologieën die zij wilden vervangen, maar ook met elkaar. Zo streden Edison en Westinghouse bij de uitrol van elektriciteit publiekelijk om de standaarden AC en DC. De makers van auto's hielden races tegen elkaar. En al eerder won bij de spoorwegen de befaamde locomotief *Rocket* een wedstrijd voor de opening van de spoorlijn tussen Liverpool en Manchester en demonstreerde daarmee voor het bredere publiek het belang van de spoorwegen.[277]

Deze technologieën waren in dit vroege stadium nog erg nieuw en onduidelijk was hoe ze gebruikt zouden worden. Dit soort publieke evenementen hielp daarom om mensen er vertrouwd mee te maken. De wedstrijden vonden echter vaak plaats in gecontroleerde omgevingen en de publieke rivaliteit ging gepaard met boude uitspraken, wat allebei bijdroeg aan de opgeklopte verwachtingen van wat de technologie in de praktijk zou kunnen. Een recent voorbeeld van de rol van publieke rivaliteiten rondom een nieuwe technologie komt uit de ruimtevaart, waar grote ondernemers als Elon Musk, Jeff Bezos en Richard Branson met elkaar wedijveren bij raketlanceringen en elkaars technologie publiekelijk bespotten.[278]

Weer een ander kanaal voor utopische verwachtingen zijn publieke tentoonstellingen. In Parijs in 1881 en bij het Crystal Palace in Londen in 1882 toonde Edison op grootse wijze de mogelijkheid van elektriciteit aan het brede publiek, wat werd gevolgd door lovende recensies in de kranten.[279] Dergelijke publieke tentoonstellingen waren tegelijkertijd ook focuspunten voor critici en activisten die bij het Crystal Palace in Londen aandacht vroegen voor de slechte werkcondities van arbeiders en betere veiligheid.[280]

---

275 Edgerton 2008: 113-114.
276 Dommering 2000: 487.
277 Freeman en Louçã 2001: 203.
278 Davenport 2019.
279 Bakker en Korsten 2021: 11
280 Van der Vleuten et al. 2017: 25.

Grote zorgen

Naast overspannen verwachtingen gaat de kennismaking met nieuwe systeemtechnologieën steevast gepaard met angstbeelden. Een terugkerend angstbeeld betreft het effect op werk. Bij breed toepasbare technologieën doemt vaak het beeld op dat de techniek de mens zal vervangen, met als gevolg dat de inkomsten van verschillende beroepsbeoefenaren wegvallen. In het verlengde hiervan ligt het beeld van het menselijk instrument dat zich tegen zijn schepper keert door diens inkomsten te ondermijnen. Dat idee komen we bij uitstek tegen bij AI maar heeft een lange geschiedenis. Jonathan Taplin laat zien hoe het internet de inkomsten van muzikanten heeft ondermijnd.[281] Die sector heeft echter een lange geschiedenis van technologische ontwrichting. Al over het effect van de platenindustrie op muzikanten zei een vakbondsleider dat nergens anders "in het mechanische tijdperk creëert de werker het instrument dat hem vernietigt, maar dat gebeurt wanneer een muzikant voor een opname speelt".[282]

Een ander dystopisch beeld is dat een nieuwe systeemtechnologie gepaard zal gaan met de teloorgang van een waardevolle manier van leven. Dat argument werd gebruikt tegen de introductie van mechanisering in de landbouw. Vanuit dat angstbeeld steunden boeren eind negentiende eeuw de Populistische Beweging in de VS.[283] Arbeiders, maar ook gevestigde industrieën, zijn vaak een bron voor dit beeld van een nieuwe technologie.

Een angstbeeld dat bij uitstek relevant is voor het heden, betreft het artificiële karakter van een nieuwe technologie. Daarmee wordt de technologie afgeschilderd als een zonde tegen de natuur of de wil van God. Dat speelt bijvoorbeeld rondom biotechnologie, maar ook elektrische straatverlichting werd gezien als verzet tegen Gods ordening van licht en duisternis. Terwijl Berlijn dus positief werd neergezet als stad van licht, werden Duitse steden in een boek van Jules Verne afgebeeld als 'Stahlstadt', steden van staal die macht en destructie symboliseren.[284] Ook een innovatie als margarine moest het ontgelden vanuit de aantijging een kunstmatige, onnatuurlijke vorm van boter te zijn en dus iets slechts.[285]

---

281 Taplin 2017.
282 Juma 2016: 213.
283 Juma 2016: 103.
284 Kaiser en Schot 2014: 192.
285 In dat geval werd dit argument ook nog eens aangevuld met het beeld van de nieuwe technologie als onpatriottisch. Toen kokosnootolie als grondstof werd gebruikt, werd margarine door tegenstanders gepresenteerd als steun voor boeren in de Filipijnen en een ondermijning van Amerikaanse boeren (Kaiser en Schot 2014: 113).

Naast angstbeelden die ontstaan op basis van die argumenten, zijn er ook angstbeelden die op meer emotioneel vlak ontstaan, door bijvoorbeeld de macht van woorden. Rondom biotechnologie bijvoorbeeld zijn termen ontstaan als 'genetische vervuiling', 'Frankenfoods' en 'Frankenfish' (voor kweekzalm).[286] We hebben bij de introductie van elektriciteit de rivaliteit tussen Edison en Westinghouse reeds benoemd. Daarbij heeft Edison heel bewust gepoogd om zijn concurrent te associëren met schrikbeelden. Hij deed experimenten met een hond om te laten zien dat de AC-standaard van Westinghouse, in tegenstelling tot zijn DC, fataal was voor het dier. Ook voerde Edison campagne om AC te gebruiken voor de elektrische stoel, om die standaard nog sterker met de dood te associëren. Als gevolg daarvan verbond een tijdschrift in 1889 de woorden 'elektro' en 'executie' om het nieuwe woord 'elektrocutie' te maken.[287] Een subtielere vorm waarmee angstbeelden aan een nieuwe technologie verbonden raken, is via geruchten. Rondom veel technologieën gingen, zeker in de beginfase, ongefundeerde claims de ronde dat ze de gezondheid zouden schaden, gevaarlijke of vuile ingrediënten zouden bevatten en zelfs tot steriliteit zouden leiden.

Er zijn dus terugkerende schrikbeelden rondom de introductie van nieuwe technologieën. Onterechte of overtrokken angsten kunnen leiden tot een algehele afkeer waardoor ook de vruchten ervan niet geplukt kunnen worden. Juma beschrijft de gelijktijdige opkomst van de mobiele telefoon en GMO. Terwijl de eerste technologie wereldwijd zonder veel verzet is ingevoerd, is de tweede in de VS omarmd en in Europa geweerd. In het geval van nucleaire technologie is de beeldvorming sterk beïnvloed door de rampen van Tsjernobyl en Fukushima, met gevolgen voor het recente beleid van veel landen.[288] Relevant in deze voorbeelden is niet zozeer of GMO en nucleaire energie meer gebruikt zouden moeten worden, maar het feit dat framing en angstbeelden in de publieke ruimte een bepalende rol kunnen spelen bij de acceptatie van een nieuwe technologie.

Slechts voor een deel kunnen angstbeelden met argumenten geadresseerd worden. Uitleg van bovenaf en technische oplossingen zijn echter veelal onvoldoende gebleken om bepaalde beelden te weerleggen, zeker als die ook drijven op de – vaak emotionele – kracht van woorden en geruchten. Als er eenmaal publiek wantrouwen richting een technologie is ontstaan – een maatschappelijke *backlash* –, is het heel moeilijk om dat te repareren. Vaak raken allerlei losstaande zaken dan met elkaar geassocieerd en problemen met elkaar verward. Daarnaast hebben angsten niet alleen betrekking op de technologie zelf, maar

286 Kaiser & Schot 2014: 309.
287 Kaiser en Schot 2014: 164-165.
288 Een voorbeeld hiervan is de ontwikkeling van beleid rondom kernenergie in Italië (Juma 2016).

kunnen ze ook voortkomen uit de indruk dat autoriteiten bij het benutten van de technologie de belangen en veiligheid van burgers onvoldoende beschermen.[289]

Voor systeemtechnologieën als AI, die de samenleving veel goeds kunnen brengen, is het zaak om dergelijke situaties te voorkomen. Aan de andere kant is het zaak om behoedzaam te zijn voor de effecten van een hype rondom een nieuwe systeemtechnologie. Deze opgave kan slechts deels door de overheid geadresseerd worden. Het is een zaak van de bredere publieke beeldvorming, waarbij een krachtenveld van onderzoekers, media, scholen en ook individuele burgers een belangrijke rol te spelen heeft. Als grote gebruiker van nieuwe technologie kan de overheid echter wel prominent invloed uitoefenen op dit proces.

Daarnaast kan de overheid ook directer beleid voeren ten aanzien van deze opgave. Dat kan via haar eigen communicatie en de voorbeeldrol in de wijze waarop ze de technologie concreet benut, maar ook door partijen te steunen die aan publieke educatie doen, zoals experts en de media. Rondom de mechanisering van de landbouw richtte de Amerikaanse overheid bijvoorbeeld instituten en groepen bij universiteiten op die de bewustwording van nieuwe technologieën in de samenleving vergrootten.[290]

Kernpunten – Opgave 1: Demystificatie

- Vanwege hun generieke karakter spreken systeemtechnologieën tot de verbeelding. Ze gaan gepaard met zowel onrealistisch hoge verwachtingen van vooruitgang als doembeelden.
- Algemeen optimisme over technologie, publieke competities en evenementen kunnen leiden tot te hooggespannen verwachtingen van een nieuwe technologie.
- Terugkerende angstbeelden bij de introductie van systeemtechnologieën gaan over het verlies van werk, de teloorgang van een manier van leven en het tegennatuurlijke karakter van de nieuwe technologie.
- Naast argumenten spelen emoties en de kracht van woorden en frames een grote rol bij angstbeelden over technologie.
- Zowel overtrokken verwachtingen als angstbeelden kunnen leiden tot afkeer van een technologie. Om de kansen van een nieuwe systeemtechnologie te kunnen benutten en tijdens het proces van maatschappelijke inbedding de juiste vragen te kunnen stellen met het oog op de mogelijke risico's, is demystificatie is daarom vereist.

289 Gezondheidsraad 2006: 111.
290 Juma 2016: 134.

3.4 Opgave 2: Contextualisering

Waar de eerste opgave speelt op het niveau van de beeldvorming, is de tweede opgave gericht op het gebruik van een systeemtechnologie. Deze opgave gaat over wat ervoor nodig is om datgene dat in het lab is ontwikkeld, toe te passen in concrete maatschappelijke praktijken. Dit is een brede opgave die verschillende dimensies kent. De opgave is ook complex, wat een van de belangrijkste redenen is waarom het zo lang duurt voordat een systeemtechnologie in de samenleving ingebed raakt. Dat iets in het lab werkt, betekent nog niet dat het ook in de praktijk functioneert. Ten aanzien van AI verschenen de laatste jaren veel nieuwsberichten over algoritmes die betere diagnoses van verschillende ziekten kunnen stellen dan artsen, de oordeelsvorming door rechters kunnen verbeteren of beter kunnen vertalen dan mensen. Dat zij vooralsnog het werk niet hebben overgenomen van artsen, rechters en tolken, heeft veel te maken met de opgave van contextualisering. Terwijl mythen inbedding kunnen verhinderen door maatschappelijk verzet of desillusie, brengt deze opgave obstakels voor inbedding in beeld die te maken hebben met het niet functioneren van de technologie in de praktijk. De leidende vraag bij deze opgave is dan ook: hoe gaat de technologie werken? (figuur 3.4)

Voor de beantwoording van die vraag gaan wij uit van een ecosysteembenadering. De opgave van contextualisering behelst dat een technologie ingebed moet worden in verschillende contexten of ecosystemen om uiteindelijk functioneel te kunnen zijn. Wij onderscheiden twee type ecosystemen: een technologisch en een sociaal ecosysteem.

Figuur 3.4 Opgave 2: Contextualisering

Het technologische ecosysteem: ondersteunende technologieën

Een nieuwe systeemtechnologie, of het nu om AI gaat of de stoommachine, elektriciteit of de verbrandingsmotor, kan technisch gezien niet op zichzelf functioneren. Ze functioneert altijd in een cluster of blok[291] van andere technologieën. We onderscheiden daarbij enerzijds *ondersteunende* en anderzijds *emergente* technologieën.

De eerste betreffen aanverwante technologieën die strikt gezien niet tot de systeemtechnologie zelf behoren, maar wel van meet af aan nodig zijn om die te kunnen laten werken. Om een verbrandingsmotor in een auto te laten functioneren is bijvoorbeeld een staalindustrie vereist. Een van de factoren die bijdroegen aan het succes van automaker Ford, was het grote netwerk van dealers en verkooppunten voor banden, batterijen en reserveonderdelen.[292] Een andere ondersteunende technologie is een op de auto aangepast wegennetwerk. In de VS waren de Federal Road Act van 1916 en de Federal Highway Act van 1921 cruciale bouwstenen om het technische ecosysteem van de auto te realiseren.[293]

Zonder dergelijke ondersteunende technologieën is een systeemtechnologie in het beste geval maar beperkt functioneel. Het wordt vanuit dit perspectief gemakkelijker om te begrijpen dat mensen toentertijd twijfelden of de auto wel echt in gebruik genomen zou worden. Zeker als deze vergeleken werd met het veel wendbaardere paard, waarvoor ht niet nodig was de omgeving aan te passen.

Hetzelfde vraagstuk speelde bij de tractor. De introductie van de tractor op het platteland was niet simpelweg de vervanging van het ene instrument door het andere, maar deze vroeg om een heel nieuwe infrastructuur van grondstoffen en toeleveranciers. De eerste tractoren waren ook niet zo betrouwbaar als paarden. Gedurende lange tijd heerste het idee dat het paard en de tractor naast elkaar zouden blijven bestaan voor verschillende doeleinden. De eerste tractoren in de VS waren niet beter dan paarden, maar werden in het westen van het land gebruikt voor grote stukken open prairie waarvoor onvoldoende paarden beschikbaar waren.[294] Pas met de tijd werd duidelijk dat de tractor het paard op het platteland zou gaan vervangen.

---

291 Alessandro Nuvolari bekritiseert de GPT-literatuur omdat die de aandacht te nauw op een enkele technologie richt. In plaats daarvan stelt hij voor te spreken van *'development blocks'* die in het geval van bijvoorbeeld ICT bestaan uit halfgeleiders, computers, software en netwerkapparatuur (Nuvolari 2019: 8).

292 Gordon 2016: 154.

293 Bakker en Korsten 2021: 17.

294 Juma 2016: 125.

Het technologische ecosysteem: emergente technologieën

Naast ondersteunende technologieën bestaat het cluster of technologische ecosysteem van een systeemtechnologie ook uit wat wij emergente technologieën noemen. Dat zijn technologieën die, anders dan ondersteunende technologieën, onafhankelijk worden ontwikkeld, maar na verloop van tijd met elkaar verbonden worden en een gezamenlijk cluster gaan vormen. Een ontwikkeling van buiten kan op deze manier een grote, niet voorziene impuls geven aan een systeemtechnologie. Neem het voorbeeld van elektriciteit. De implementatie van deze technologie in de huishoudens verliep aanvankelijk moeizaam, mede omdat er voor het creëren van licht simpeler manieren bestonden, zoals kaarsen en gaslampen. Doordat de ontwikkeling van huishoudelijke apparatuur, zoals het strijkijzer, en later ook allerlei elektronica nieuwe toepassingen van elektriciteit mogelijk maakte, kreeg het gebruik ervan echter een grote impuls.

De barcode is ook een innovatie die pas na contextuele aanpassingen goed ging werken. De eerste barcodescan vond plaats in het midden van de jaren zeventig. Het duurde echter dertig jaar voordat organisaties in de gehele productieketen de complementaire technologische, organisatorische en procesveranderingen hadden toegepast en het gebruik van de barcode gangbaar werd.[295]

Een recenter voorbeeld is de opkomst van e-commerce. Over die consumententoepassing bestaan al hoge verwachtingen sinds de uitrol van het internet. Het bedrijf Amazon is in 1994 opgericht en was in de jaren negentig een van de gehypte bedrijven in de 'dotcombubbel' vanuit de belofte dat mensen via het internet hun aankopen zouden gaan doen. Ook na de crash van 2000 heeft het bedrijf nog jarenlang in e-commerce geïnvesteerd zonder eraan te verdienen. Het heeft meer dan twee decennia geduurd voordat het online boodschappen doen een vlucht nam. Complementaire innovaties, zoals een veilige en gemakkelijke manier van onlinebetalingen en een betere logistieke infrastructuur zoals regionale sorteercentra, lagen hieraan ten grondslag. Hetzelfde geldt voor transportdiensten als Uber, SnappCar en Greenwheels. Het idee van het online regelen van taxiritten en carpoolen bestaat al decennia, maar ook deze diensten nemen de laatste jaren pas een vlucht. Dat heeft te maken met de opkomst van technologieën zoals GPS in mobiele telefoons, waardoor de dienst lokaal aangeboden kan worden.

Door de noodzakelijke ontwikkeling van een heel technisch ecosysteem van ondersteunende en emergente technologieën duurt het lang voordat een nieuwe systeemtechnologie ook daadwerkelijk optimaal gaat functioneren. Het traject ervan wordt daardoor ook onvoorspelbaar: de technologie zelf

295 Pethokoukis, 25 november 2019.

maakt verbeteringen door, er vinden complementaire innovaties plaats, prijzen dalen[296] en nieuwe systemen en toepassingen worden ontwikkeld. Zelfs als een systeemtechnologie in eerste instantie geen voet aan de grond lijkt te krijgen, kunnen dergelijke veranderingen onder de radar plaatsvinden waardoor de nieuwe technologie plotseling superieur wordt aan de oude methoden en het gebruik ervan momentum krijgt.

Enveloping

Voor de contextualisering van AI in de technologische context is '*enveloping*' een relevant concept: het creëren van een omgeving waarbinnen een technologie tot haar recht komt. De eerder aangehaalde Luciano Floridi, die het concept populariseerde in het kader van AI, bestrijdt het idee dat technologie een instrument is. Dat suggereert namelijk een heel oud model van een menselijke gebruiker die door middel van een technologie een effect heeft op een natuurlijke omgeving. Dat model past op een speer, een bijl of een paraplu, maar veel technologieën werken niet als een instrument in op een natuurlijke omgeving (respectievelijk de prooi, de boom of het zonlicht). Zij werken in op andere technologieën. Dat gold al voor de hamer (met de spijker), maar bij uitstek voor alle technologie sinds de Industriële Revolutie. Een auto kan niet goed in de natuur rijden, maar wel als de omgeving wordt aangepast met een verharde weg. Dat proces waarbij wij de omgeving van een technologie aanpassen om die beter te laten functioneren, heet *enveloping*.[297] Het inzicht dat wij hiervan meenemen, is dat het gebruik van een technologie niet alleen gestimuleerd wordt door verbetering van die technologie zelf, maar ook door aanpassing van de omgeving.

Enveloping roept de vraag op wie zich waaraan aanpast: de technologie aan de mens of de mens aan de technologie. Alhoewel dit laatste verzet oproept, is het belangrijk om ons te realiseren hoe gangbaar dat is. We hoeven alleen een blik op de straat te werpen om te zien hoe wij onze omgeving met geasfalteerde wegen, parkeerplaatsen, verkeersborden en verkeersregels in hoge mate aan de auto hebben aangepast. Dergelijke dynamieken zullen we bij AI veelvuldig tegenkomen.

Het sociale ecosysteem: macro-economische context

Naast het technologische ecosysteem gaat contextualisering ook over inpassing in het sociale ecosysteem. Een eerste onderdeel daarvan is de macro-economische context. Een nieuwe technologie heeft namelijk een eigen logica die vanuit

296 Prijsdalingen zijn ook van groot belang om het gebruik van een technologie in de praktijk te laten werken. Een studie laat bijvoorbeeld zien hoe, vergeleken met het begin van de negentiende eeuw, de prijs van licht met een factor vierhonderd is gedaald (Agrawal et al. 2018: 11).

297 Floridi 2014: 144.

zichzelf niet meteen aansluit op bestaande processen in organisaties. Dat gaat niet over één nacht ijs. Organisaties hebben vaste manieren van werken en dat maakt het moeilijk om nieuwe benaderingen uit te proberen.

Voor organisaties in gevestigde industrieën geldt dan ook vaak "de vloek van kennis".[298] Het simpelweg aanschaffen van nieuwe machines of zelfs het opzetten van een afdeling ervoor, zoals een IT-afdeling of een AI-afdeling, is niet afdoende. Hedendaagse organisaties hebben immers ook geen elektriciteitsafdelingen meer. Die gevestigde systeemtechnologie is namelijk door het hele proces van organisaties heen ingepast. Dat heeft enige tijd gekost. Fabrieken moesten bijvoorbeeld reorganiseren voor de aanleg van elektriciteitskabels.[299] Ook de opkomst van de telefoon en de typemachine droegen uiteindelijk bij aan de mechanisering en bureaucratisering van het kantoor, en daarmee aan het groeien van organisaties, maar het kostte tijd om die transformatie door te voeren.[300]

Naast tijd kost het vaak ook veel kapitaal om erachter te komen hoe zo'n transformatie moet gebeuren. Dit werpt licht op het fenomeen van de productiviteitsparadox bij nieuwe systeemtechnologieën. Het duurde jaren voordat elektriciteit de economie netto productiever maakte.[301] Een van de verklaringen van de aanvankelijk achterblijvende productiviteit heeft te maken met de toelevering van energie. Zo werd de stoommachine in Engeland aanvankelijk alleen toegepast op plaatsen nabij koolmijnen.[302] Om systeemtechnologieën productief te maken is het dus van belang om aandacht te hebben voor de bredere organisatie van processen waarbinnen deze moeten gaan werken.

Het sociale ecosysteem: gedragscontext

Het sociale ecosysteem gaat naast deze (macro-)economische context ook over inbedding in de gedragsmatige context. Dat geldt zowel voor consumenten als voor de gebruikers die in organisaties met de nieuwe technologie aan de slag moeten. Die laatste groep moet vaak opgeleid en getraind worden in het gebruik van de toepassingen die de technologie faciliteert. Hieronder valt dus het grotere vraagstuk van aanpassingen op de arbeidsmarkt.[303] Terwijl in de labfase fundamentele kennis van een technologie vereist is, verschuift gaandeweg de nadruk naar toepassingskennis in verschillende domeinen. Ten tijde van de

298 Brynjolfsson et al. 2019: 42.
299 Bakker 2017.
300 Freeman en Louçã 2001: 28.
301 Agrawal et al. 2019.
302 Baker en Korsten 2021: 6-7.
303 Zie het WRR-rapport *Het Betere Werk* over technologisering van werk (WRR 2020).

inbedding van elektriciteit vroeg dit om tal van ingenieurs en uitvinders die bedachten in wat voor contexten elektriciteit effectief gebruikt kon worden.

Mensen die een nieuwe technologie gaan gebruiken, moeten er vertrouwen in krijgen, tot op zekere hoogte begrijpen hoe deze werkt en ze moeten ermee willen werken. Daarvoor moeten de juiste prikkels aanwezig zijn, alsook oog voor prikkels die inbedding juist tegenwerken. Mensen zullen een nieuwe technologie niet omarmen als ze daarmee hun eigen werk overbodig maken of hun eigen verdienmodel ondermijnen. Denk aan de artiesten in de opnamestudio's voordat een nieuw verdienmodel was ontwikkeld rondom onlinestreamingdiensten. Daarnaast zullen professionals zoals artsen, rechters of accountants een nieuwe technologie ook niet accepteren dan wel ten volle benutten als de technologie (nog) niet in staat is om de standaarden van de betreffende professie invulling te geven.[304]

Ook van consumenten vraagt een nieuwe technologie vaak een gedragsverandering. Neem wederom het voorbeeld van de muziekopnames. Voordat die bestonden, luisterden mensen alleen op specifieke gelegenheden bij een liveoptreden naar muziek. Radio's en platen maakten geheel nieuwe praktijken van muziek luisteren in het eigen huis mogelijk, maar consumenten moesten daar wel aan wennen.

Net als demystificatie is contextualisering een brede opgave waar de overheid maar beperkt grip op heeft. Een groot deel van de contextualisering van een technologie gebeurt in de vele duizenden contexten waarin mensen op de werkvloer aan de slag gaan met een nieuwe technologie en leren wanneer deze gaat werken. Dat is een iteratief proces. Wel heeft de overheid verschillende mogelijkheden om de brede opgave van contextualisering te faciliteren en richting te geven.

Zo kan de overheid bijdragen aan contextualisering door te investeren in ondersteunende en emergente technologieën. Dat deed de Amerikaanse overheid voor de auto door autowegen aan te leggen. Een tweede manier is om zelf aan dat proces van experimenteren in contexten mee te doen. Als gebruiker van een nieuwe technologie draagt de overheid bij aan de vorming van een markt, kan zij hoge standaarden stellen en kan zij gebruikmaken van een voorbeeldfunctie richting de private sector. De overheid kan ook bijdragen via haar aanbestedingsbeleid, omdat zij als grote speler een markt kan stimuleren.

304 Van Ettekoven en Prins 2018.

Kernpunten – Opgave 2: Contextualisering

- Om een technologie te laten werken is contextualisering nodig. Daarbij gaat het over het begrijpen en benaderen van de technologie in bredere socio-technische ecosystemen.
- Het technische ecosysteem bestaat enerzijds uit ondersteunende technologieën die de werking van een systeemtechnologie mogelijk maken.
- Anderzijds bestaat het uit emergente technologieën: geheel andere technologieën die onafhankelijk opkomen, maar een onverwacht sterke impuls aan een technologie kunnen geven.
- Een relevant proces bij de contextualisering van systeemtechnologieën is 'enveloping': het aanpassing van de omgeving aan een technologie.
- Het sociale ecosysteem bestaat ten eerste uit de macro-economische context en betreft complexe vraagstukken over productiviteit en de organisatie van werkprocessen.
- Daarnaast betreft het de gedragsmatige context, met de prikkels, praktijken, standaarden en overtuigingen van mensen die met de technologie te maken hebben.

3.5 Opgave 3: Engagement

De eerste opgave speelt op het niveau van de beeldvorming en de tweede op dat van het gebruik. De derde opgave, engagement, bevindt zich op het niveau van de maatschappelijke omgeving. De opgave gaat over de mensen die door de systeemtechnologie geraakt worden en de partijen die er om die reden bij betrokken (moeten) zijn (figuur 3.5). Naast techneuten gaat het dan ook om burgers en het maatschappelijk middenveld.

Figuur 3.5 Opgave 3: Engagement

Waarden, belangen en idealen

Zoals gezegd, hangen de opgaven sterk met elkaar samen. Bij de vorige opgave kwamen we de menselijke omgeving al tegen bij het sociale ecosysteem. De centrale vraag daar was hoe de technologie te laten werken. De opgave van engagement gaat daarentegen over de betrokkenheid van mensen bij de vormgeving en inzet van die technologie. Dat is van belang omdat zij met hun waarden, belangen en idealen bijdragen aan de inbedding van een nieuwe systeemtechnologie. Uiteraard kan rekening houden met de belangen van mensen ook bijdragen aan de functionaliteit van een technologie, maar het uitgangspunt van deze opgave is dat het voor de maatschappelijke inbedding op de langere termijn inherent van belang is dat verschillende groepen mensen betrokken zijn bij dit proces. Zo wordt technologie 'gehumaniseerd' of 'gedemocratiseerd'.

Aandacht voor deze opgave is in het bijzonder van belang gebleken in de fase waarin een technologie het lab verlaat, omdat nog verkend moet worden onder welke voorwaarden deze technologie aan de samenleving bijdraagt. Het betrekken van het maatschappelijk middenveld is bovendien cruciaal omdat elke technologie verbonden is met machtsverhoudingen. Het zijn veelal sterke partijen als grote bedrijven en overheden die een nieuwe technologie als eerste gebruiken. Zo helpt een nieuwe technologie in eerste instantie vaak om bestaande machtsverhoudingen te versterken. Engagement is nodig zodat ook andere partijen uit de samenleving een stem krijgen bij de wijze waarop de nieuwe technologie ingezet wordt.

Een spectrum van betrokkenheid

Engagement kan verschillende vormen aannemen. Aan het ene eind van het spectrum kunnen groepen zich sterk verzetten tegen het gebruik van bepaalde technologieën en een verbod daarop nastreven. Dit verzet kan zelfs gepaard gaan met geweld. Aan het andere eind van het spectrum kan engagement een vorm van meedenken zijn, waarbij partijen hun expertise, waarden en verlangen inbrengen om de technologie anders te gebruiken. Op die manier kunnen belanghebbenden direct een ander gebruik ontwikkelen.

Ze kunnen dat ook op indirecte wijze doen door overheden op te roepen tot regulering, de vierde opgave die wij hierna zullen behandelen (paragraaf 3.6). Hier is van belang dat geëngageerde partijen uit de samenleving, door zich te mobiliseren, technologie kunnen helpen reguleren. Ook daarvoor is betrokkenheid juist in een vroege fase van een technologie van belang. Onzekerheid over de ontwikkeling ervan maakt het namelijk moeilijk voor overheden om te weten hoe zij de technologie kunnen reguleren. Partijen uit het maatschappelijk middenveld vervullen dan een belangrijke signalerende en delibererende functie voor de politiek en overheid. De hoofdvraag bij deze opgave is kortom: wie moeten er betrokken zijn?

Winnaars en verliezers

Individuele mensen of belangengroepen raken om verschillende redenen geëngageerd bij de inbedding van een systeemtechnologie in de samenleving. Vaak hebben die redenen te maken met winnaars en verliezers. Nieuwe technologieën brengen immers voordelen voor de samenleving, maar creëren naast winnaars ook altijd verliezers. Toen Schumpeter over creatieve destructie sprak, zag hij ook de ellende die nieuwe technologie veroorzaakte en visualiseerde hij grote delen van de samenleving die verpletterd werden onder 'de wielen van innovatie'.[305] Naast bedreiging van banen gaat het proces van innovatie en experimenteren vaak samen met ongevallen en zelfs roekeloos en gevaarlijk gedrag. We noemden eerder de praktijken rondom de massaproductie van melk en de introductie van margarine. Vaak gebruikten producenten allerlei chemicaliën voor kleur en houdbaarheid die schadelijk waren voor de gezondheid van burgers.[306] Regelmatig benadelen technologieën ook specifieke categorieën binnen het maatschappelijk middenveld, zoals consumenten of werknemers. Veelal zijn dit partijen met een zwakkere of afhankelijke positie, omdat sterke partijen de technologie als eerste voor hun doelen weten in te zetten dan wel hun expertise en positie als *first mover* in hun voordeel laten werken. Zo vergroten nieuwe systeemtechnologieën initieel bestaande machtsongelijkheden.

De stoommachine ging gepaard met angst voor de marginalisering van de arbeidersklasse. Bij de opkomst van de spoorwegen waren rijkere passagiers bezorgd over het contact met armeren in de trein, waardoor het systeem van meerdere klassencoupés ontstond.[307] Elektriciteit in de vorm van straatverlichting werd gezien als een vergroting van de macht van de overheid ten opzichte van burgers. Bij de auto speelden klassenverschillen eveneens een rol: de auto werd gezien als een instrument van de welvarende witte elite die gaandeweg andere burgers van de straat verjoeg.[308]

Als gevolg van deze verhoudingen raakten groepen burgers met de nieuwe technologie geëngageerd. Een van de vormen die dat engagement kon aannemen was protest, waar in extreme gevallen ook geweld bij werd gebruikt. In de jaren tien van de negentiende eeuw kwamen in Engeland fabriekswerkers, de Luddieten, in opstand tegen de mechanisering van de arbeid en zij vernielden machines. Tijdens de Plug Riots in 1842 staakte een half miljoen arbeiders, die ook stoommachines onklaar maakten. Dit was een van de weinige instrumenten die arbeiders hadden, omdat de Britse overheid in die tijd erg weinig

305 Schubert 2013.
306 Juma 2016: 97.
307 Van der Vleuten et al. 2017: 47.
308 Bakker en Korsten 2021: 30.

deed om arbeiders te beschermen.[309] De term 'Luddiet' wordt tegenwoordig vaak gebruikt voor mensen die de futiele poging doen om technologische vooruitgang tegen te gaan. Zo simpel is het echter niet. Met hun rellen wisten de Luddieten de mechanisering van de textielindustrie wel degelijk af te remmen en creëerden zij solidariteit tussen arbeiders, wat later de basis zou vormen voor de vakbonden.[310] Het doel was dus niet zonder meer om een nieuwe technologie af te wijzen, maar veeleer om op te komen voor de positie van werknemers.

Ook de introductie van de auto ging gepaard met protest door benadeelde zwakkere partijen. Daarbij ging het om de gevaren die duidelijk werden nadat de eerste dodelijke ongelukken hadden plaatsgevonden, maar vooral ook de 'strijd om de straat'. Marktkooplui, paarden en voetgangers werden namelijk gaandeweg van de weg gedrukt. Paarden werden gezien als een bron van congestie en omgekeerd werd geklaagd over de ruimte die autoparkeerplaatsen innamen. In de loop van de jaren dertig slaagde de autolobby erin de bevolking ervan te overtuigen dat de wegen vooral bedoeld waren voor auto's. Dat gebeurde middels educatie, waarbij kinderen geleerd werd om op te passen bij het oversteken. Naast de regels voor auto's werd ook een hele reeks van geboden en verboden voor fietsers en voetgangers opgesteld, zoals het strafbaar stellen van schuin oversteken. De autolobby pleitte voor snelle banen waar alleen auto's mochten komen, waardoor de snelwegen ontstonden. Er was met andere woorden een strijd over wie, onder welke voorwaarden, de legitieme gebruiker van de straat is en wie niet.[311]

Een recenter voorbeeld van protest is de anti-kernenergiebeweging. Die gebruikte posters, krantenberichten en stickers, hield demonstraties zoals *'die-ins'* en vormde menselijke ketens. Soms ook saboteerde ze apparatuur.[312] Dit soort protesten heeft uiteindelijk bijgedragen aan de bredere maatschappelijke en politieke discussie over kernenergie.

Roep om regels

Zoals het voorbeeld van de auto laat zien is, kon het engagement van het maatschappelijk middenveld ook de vorm aannemen van campagnes richting de politiek om beleid of regels te maken. Halverwege de negentiende eeuw droeg de beweging van *Chartism* in Engeland bij aan wetgeving die de maximale werkuren voor jonge mensen en vrouwen reduceerde.[313] In de VS voerden vrouwenorganisaties aan het eind van de negentiende eeuw campagne voor

309 Bakker en Korsten 2021: 9.
310 Juma 2016: 26-27.
311 Van der Vleuten et al. 2017: 84-86.
312 Van der Vleuten et al. 2017: 135.
313 Freeman en Louçã 2001: 172-173.

betere werkcondities. De Woman's Christian Temperance Union (WCTU) streed tegen alcohol, maar ook tegen het brede gebruik van allerlei nieuwe medicijnen. Zo droegen ze met hun campagnes bij aan wetgeving gericht op het labelen van ingrediënten en op grenzen aan de vrije distributie van medicijnen.[314] Om het autoverkeer veiliger te maken bedacht een activistische ingenieur uit Delft in 1970 de verkeersdrempel en een paar jaar later kondigde de Nederlandse overheid het concept van woonerven aan, waardoor zones werden gecreëerd waar voetgangers weer voorrang hadden op auto's.[315]

Partijen uit het maatschappelijk middenveld konden het gebruik van een nieuwe technologie ook op directe wijze beïnvloeden, dus niet via de politiek. Dat kon door de technologie zelf in te zetten voor eigen doelen. Bellamy-clubs in de VS organiseerden zich om technologieën te gebruiken voor utopische gemeenschapsdoelen. Vakbonden, feministen, artsen en voedingsspecialisten zetten zich in om de technieken en apparaten van het moderne huis gezond, veilig en prettig te maken. Gebruikersgemeenschappen ontwierpen zelfs appartementenblokken met gemeenschappelijke ruimten voor koken en kinderopvang om zo gemeenschap en gelijkheid te stimuleren. Bij de telefoon gebruikten vrouwen en migranten dit apparaat anders dan de telefoonmaatschappijen hadden bedoeld, wat er uiteindelijk toe leidde dat de diensten werden aangepast.[316]

Rondom de massaproductie van kleding slaagde de White Label League erin om producenten ervan te overtuigen een wit label toe te voegen als de beweging de werkcondities had goedgekeurd.[317] Op het gebied van digitalisering is het project Claudette een mooi voorbeeld. Daarin wordt geprobeerd de positie van consumenten te versterken door van talloze onlineplatforms de rechtmatigheid van hun servicevoorwaarden geautomatiseerd te beoordelen en voor consumenten inzichtelijk te maken.[318]

Opkomen voor publieke belangen

Burgers die met nieuwe technologieën te maken kregen, raakten daar dus op allerlei manieren bij betrokken: ze ondervonden de effecten, gaven het gebruik richting en lieten hun stem horen. Een aantal maatschappelijke partijen speelt in het bijzonder een belangrijke rol. Een daarvan is de media, die we ook tegenkwamen bij de opgave van demystificatie. Waar het bij die laatste opgave gaat om de rol van het informeren van het publiek, gaat het hier om het aankaarten van vraagstukken rondom publieke belangen.

---

314 Gordon 2016: 221-224.
315 Van der Vleuten et al. 2017: 153.
316 Van der Vleuten et al. 2017: 44-46.
317 Van der Vleuten et al. 2017: 50-51.
318 Leeuw 2020: 132-133.

Een historisch voorbeeld van dat soort mobilisatie speelt zich af rondom de introductie van elektriciteitskabels in steden. In oktober 1889 werd John E.H. Feeks, een werknemer van Western Union, in een gruwelijk ongeluk door die kabels geëlektrocuteerd. Zijn lichaam hing vijfenveertig minuten te roken en vonken totdat het naar beneden gehaald kon worden. Er ontstond breed verzet in New York en kranten vertelden verhalen van burgers die overal elektriciteitskabels doorknipten. De publieke opinie liet zien dat ondernemingen hun eigen winst boven de veiligheid van de bevolking plaatsten. *The New York Times* stelde dat de bevolking geen genoegen hoefde te nemen met egoïstische ondernemers en onwetende en corrupte ambtenaren. Deze oproer leidde tot een groter onderzoek naar de macht van dominante bedrijven en zelfs tot nieuwe modellen waarin gemeenten meer zeggenschap kregen en ook burgerparticipatie belangrijker werd.[319]

Een voorbeeld uit een heel ander domein is de opkomst van de koelingsindustrie. In plaats van met natuurlijk ijs konden goederen daardoor op artificiële wijze in warenhuizen gekoeld worden. Het publiek werd op een gegeven moment sceptisch over de macht van de '*ice trust*'. Agitatie, teweeggebracht door de kranten, leidde vervolgens tot publieke oproer, waarna het verplicht werd om producten te labelen met de datum waarop ze in de koeling waren gezet.[320] Daarnaast waren de deuren van koelkasten en vriezers aanvankelijk moeilijk te openen, waardoor spelende kinderen erin vast konden komen te zitten en verstikken. Verontwaardiging in de media hierover leidde tot de introductie van een ander type deuren.[321]

Naast de media vormen wetenschappers en andere experts een tweede belangrijke groep van een geëngageerd maatschappelijk middenveld. Zij spelen hun rol onder andere door in hun publicaties de bewustwording van mensen te vergroten en problemen en misstanden aan te kaarten. Een beroemd voorbeeld hiervan is de biologe Rachel Carson wiens boek *Silent Spring* uit 1962 een belangrijke impuls gaf aan de ecologische beweging. Haar analyse legde de keerzijde van industriële productie en landbouw bloot en mobiliseerde daarmee een beweging tegen machtige bedrijven.

Rondom de auto droeg het werk van critici als Guy Debord, Constant Nieuwenhuys, Jane Jacobs en Lewis Mumford bij aan de bewustwording van misstanden.[322] Ook kunstenaars en schrijvers van fictie kunnen bijdragen aan

319 Juma 2016: 165-166, 172.
320 Juma 2016: 185.
321 Juma 2016: 186.
322 Van der Vleuten et al. 2017: 120-121.

het engagement van het publiek. De eerdergenoemde Bellamy-clubs waren geïnspireerd door het boek *Utopia: Looking Backward* van Edward Bellamy. Ook beroemde schrijvers als H.G. Wells en Mark Twain vroegen in hun werk aandacht voor de invloed van technologieën als elektriciteit.[323] In 1906 publiceerde Upton Sinclair *The Jungle*, een boek over de verschrikkelijke condities in de vleesindustrie in Chicago. Het leidde tot een onmiddellijke reductie van de vleesconsumptie en de publieke onrust droeg eraan bij dat een systeem van inspecteurs ontstond.[324]

Een nog ouder voorbeeld ten tijde van de Industriële Revolutie was een rapport waarin een commissie van artsen aantoonde dat de burgers van Manchester daadwerkelijk ziek werden van de rook in de stad.[325] Het duurde echter nog lang voordat daar wat mee gebeurde. Zoals in paragraaf 3.1 aangegeven, is de mate van organisatie, en daarmee de mate van invloed, van deze partijen in de loop van de tijd gegroeid. Beroepsgroepen, associaties en commissies van (wetenschappelijke) experts zijn een steeds grotere rol gaan spelen bij de inbedding van een nieuwe technologie. Spraakmakende wetenschappelijke vakbladen zijn een belangrijk medium hiervoor, maar ook oproepen en bijeenkomsten. In 1955 verscheen een manifest waarin de filosoof Bertrand Russell en de natuurkundige Albert Einstein opriepen tot vreedzame oplossingen voor internationale conflicten, waar de wetenschappelijke wereld aan moest bijdragen. Daarop volgde een serie expertconferenties onder de naam 'Pugwash Conferences on Science and World Affairs'.[326]

Nadat in 1973 de technologie voor het klonen van genen was ontwikkeld, werd op de Asilomar Conference on Recombinant DNA in 1975 een vrijwillig moratorium op genetische modificatie ingesteld, zodat medische instituten ondertussen richtlijnen voor veiligheid konden ontwikkelen. Daarmee werd de basis gelegd voor een op wetenschap gebaseerd systeem van risicoanalyse.[327] Een ander voorbeeld van de invloed van experts op technologische ontwikkeling is de IPCC op het gebied van klimaat, waarvan de leden vooraanstaande wetenschappers zijn. Wetenschappers, andere experts, schrijvers en kunstenaars, maar ook burgers en hun belangenbehartigers kunnen dus campagnes voeren tegen het gebruik van nieuwe technologieën, maar zij dragen vooral bij aan manieren om die technologieën verantwoorder te gebruiken en kunnen zo het gebruik juist te stimuleren.

323 Freeman en Louçã 2001: 232.
324 Gordon 2016: 82.
325 Bakker en Korsten 2021: 9.
326 Van der Vleuten et al. 2017: 102.
327 Juma 2016: 236-237.

Een laatste observatie is dat door hun professionele nadruk op de openbaarheid van publicaties en kennis, wetenschappers en onderzoekers tegenover overheden en bedrijven kunnen komen te staan als die laatste gebaat zijn bij geheimhouding. Het Human Genome Project was een internationale samenwerking om het menselijke genoom publiek beschikbaar te maken. Tegelijkertijd had het bedrijf Celera echter een project om het genoom privaat op te slaan, waardoor het in conflict kwam met de wetenschappelijke gemeenschap. Wetenschappers en bedrijven botsten ook met elkaar over de vraag of delen van genen octrooieerbaar waren of niet.[328]

In het veld van de cryptografie botsten wetenschappers met het beleid van geheimhouding van zowel overheden als bedrijven. Wetgeving tegen de export van kennis maakte het in de jaren negentig bovendien lastig voor wetenschappers om te weten wat zij wel en niet mochten doceren aan hun buitenlandse studenten. In verzet tegen de drang van de Amerikaanse overheid om encryptiesoftware geheim te houden, maakte de programmeur Philip Zimmerman zijn programma openbaar (*open source)*. Vervolgens werd hij daarvoor aangeklaagd.[329]

De open source-beweging is een belangrijke groep van experts uit het maatschappelijk middenveld op het gebied van de digitale technologie. Tal van rechtszaken getuigen van de spanning rondom openbaarheid. Zo werd in eigen land rondom de Mifare Classic chip, onder meer gebruikt in de OV-chipkaart, een rechtszaak aangespannen tegen de wetenschapper Bart Jacobs en de Radboud Universiteit. De rechter wees de eis tot een publicatieverbod echter af.[330] Relevant voor ons betoog hier is de overweging van de voorzieningenrechter dat "er in een democratische samenleving grote belangen zijn gemoeid met het kunnen publiceren van de resultaten van wetenschappelijk onderzoek en het informeren van de samenleving over de tekortkomingen van een product, zodat maatregelen kunnen worden genomen tegen de risico's ervan."[331] Rondom de publicatie van een paper waarin de auteurs uitlegden hoe een gevaarlijke variant van de pokken was ontwikkeld, speelde een vergelijkbare spanning.[332] Als werknemers van bedrijven spelen experts niet alleen een rol bij het vraagstuk van openbaarheid, maar ook bij allerlei andere ethische vraagstukken binnen bedrijven. Na de Tweede Wereldoorlog stelde een Duitse ingenieursorganisatie bijvoorbeeld een eed in om niet te werken voor bedrijven die mensenrechten schenden.[333]

328 Huys et al. 2011: 1104-1107.
329 Leung 2019: 202.
330 Rechtbank Arnhem, 18 juni 2008.
331 Voorzieningenrechter Arnhem, 18 juli 2008.
332 Leung 2019: 150-154.
333 Van der Vleuten et al. 2017: 127.

Kernpunten – Opgave 3: Engagement

- Engagement van het maatschappelijk middenveld is belangrijk om bij het gebruik van een nieuwe technologie relevante waarden en belangen te kunnen agenderen.
- Het maatschappelijk middenveld speelt een belangrijke rol via een breed spectrum van engagementsvormen: van verzet en protest tot aan campagne voeren en het stimuleren van alternatief gebruik en ontwerp.
- De media en journalistiek zijn belangrijk om misstanden te agenderen en de publieke opinie te mobiliseren.
- Wetenschappers en andere experts kunnen onder andere standaarden en principes van goed gebruik opstellen, een cultuur van openheid over de technologie stimuleren en de nieuwe technologie benutten indachtig publieke waarden.

3.6 Opgave 4: Regulering

De vierde opgave speelt op het niveau van de samenleving als geheel en betreft de regulering van nieuwe technologie. Daarbij kan het gaan om wet- en regelgeving, maar ook om normen en standaarden. Deze opgave gaat over de vraag welke kaders nodig zijn (figuur 3.6). Bij deze opgave heeft de overheid, zowel nationaal als internationaal, bij uitstek een rol te spelen. Andere spelers zijn hierbij echter ook van belang.

Figuur 3.6 Opgave 4: Regulering

Het Collingridge-dilemma

Het opstellen van spelregels voor zoiets groots, complex en veelzijdigs als een systeemtechnologie gaat gepaard met uitdagingen, problemen en dilemma's. Welbekend in dit verband is het zogenoemde Collingridge-dilemma. Dit stelt dat het in een vroege fase lastig is om een technologie te reguleren, omdat nog veel onduidelijk is over de werking en effecten ervan. Bovendien wordt de noodzaak om te reguleren in die fase nog niet zo sterk gevoeld. In een latere fase wordt duidelijker waar en waarom regulering noodzakelijk is. De effecten van de technologie binnen de samenleving zijn op dat moment sterker zichtbaar. Veel beslissingen die in een eerdere fase zijn genomen, zijn dan echter nog maar moeilijk terug te draaien. Bovendien zijn er allerlei machtsstructuren ontstaan die niet eenvoudig en snel te veranderen zijn. In de eerste fase gaat het vooral om een informatie- en kennisprobleem en in de tweede fase om een machtsprobleem.

Een illustratie van dit dilemma betreft de architectuur van het internet, die ontwikkeld werd vanuit een geest van openheid en vrije markt. In de huidige tijd wordt echter duidelijk dat destijds allerlei zaken rondom veiligheid en beveiliging onvoldoende geregeld zijn, wat ons nu kwetsbaar maakt voor bijvoorbeeld digitale ontwrichting.[334] Om die fouten optimaal te corrigeren zouden grote delen van het hele internet opnieuw ingericht moeten worden. En dat is een immense, zo niet onmogelijke opgave.

Machtsconcentratie

Wanneer een nieuwe technologie dus wijdverspreid is, wat wij maatschappelijke inbedding noemen, is het lastig om grote veranderingen teweeg te brengen. De noodzaak daartoe groeit echter wel. Zoals bij de vorige opgave duidelijk werd, komen de eerste tekenen dat verandering noodzakelijk is van acute vraagstukken die vaak door het maatschappelijk middenveld op de agenda worden gezet. Het gaat dan om ongelukken, misbruik, opportunistisch gebruik en gevaarlijke praktijken. Gaandeweg wordt duidelijk dat er een structurelere opgave van regulering nodig is om de technologie en de impact ervan op de samenleving in goede banen te leiden. Centraal bij regulering staat dus een verbreding van de aandacht van uitsluitend acute vraagstukken naar de structurelere problematiek.

Een terugkerend voorbeeld van een structureel vraagstuk dat historisch rondom systeemtechnologieën opkomt, is machtsconcentratie. Als gevolg van de dynamiek en innovatie rondom een nieuwe systeemtechnologie ontstaan namelijk enorme machtsconcentraties bij bepaalde bedrijven die monopolistische of

334 WRR 2019.

oligopolistische macht verkrijgen. Behalve dat zo'n machtsconcentratie tot economische problemen leidt, krijgen deze bedrijven ook buitenproportioneel veel macht over de samenleving en komen publieke waarden onder druk te staan.[335] Terwijl deze grote bedrijven initieel nog worden gezien als wonderen van innovatie en brengers van maatschappelijk goed, ontstaat gaandeweg een negatiever beeld van deze bedrijven, omdat met de verspreiding van een technologie ook hun machtspositie groeit.

Met de komst van de spoorwegen ontstonden in de VS bijvoorbeeld grote zakenimperia van mensen als Andrew Carnegie en Jay Gould. Hun macht en negatieve invloed blijkt duidelijk uit de bijnaam die zij kregen: de '*robber barons*'. Ook bij elektriciteit ontstond een immense machtsconcentratie. In 1894 fuseerde de Edison Company met Thomson-Houston tot de gigant GE. Samen met Westinghouse domineerde het bedrijf de Amerikaanse markt. In Europa ontstond Siemens in Berlijn en Ganz in Boedapest. Zij behoorden tot de vroegste echte multinationals. Vlak voor de Eerste Wereldoorlog waren GE en Westinghouse uit de VS en Siemens en AEG uit Europa de grootste bedrijven van de wereld. Er bestond in die tijd veel angst voor dit 'wereldkartel'. In 1903 kwam de bestuurder Rathenau van AEG met GE daadwerkelijk tot een overeenkomst om de wereld te verdelen.[336]

Rond diezelfde tijd ontstond in de opkomende olie-industrie de gigant Standard Oil van John D. Rockefeller. Even later ontstond rondom de verbrandingsmotor een enorme machtsconcentratie in de auto-industrie, toen zich in Detroit – het Silicon Valley van die tijd – de zogenoemde Big Three vestigden: General Motors, Ford en Chrysler. In de jaren twintig waren de VS en Canada verantwoordelijk voor bijna 90 procent van de wereldwijde productie van trucks, auto's en tractoren.[337] Jarenlang zou Detroit de industrie, ook buiten de grenzen, domineren. De uitspraak "wat goed is voor General Motors is goed voor Amerika en vice versa", toegeschreven aan Charles Erwin Wilson, is typerend voor de invloed die dit bedrijf op het land had. Halverwege de twintigste eeuw was AT&T de gigant van de telecommunicatie en het bijbehorende Bell Labs was wereldwijd een motor van innovatie. In de computerindustrie ontstond het machtige IBM. De filmklassieker *2001: A Space Odyssey* verbeeldt de gevaarlijke kant van dit bedrijf. De kwalijke computerintelligentie in die film heet HAL, een naam die ontstaat als je de drie letters van IBM één letter in het alfabet naar voren haalt.

335 Prüfer en Schottmüller 2017.
336 Freeman en Louçã 2001: 244.
337 Freeman en Louçã 2001: 260.

Naarmate machtsconcentratie een groter maatschappelijk vraagstuk wordt, zien we een historisch patroon dat machtige bedrijven campagne voeren om het oplossen van problemen aan de markt of aan zelfregulering over te laten. Vaak was dat ook een manier om dwingende externe regels af te houden. *Robber barons* als J.D. Rockefeller en J.P. Morgan presenteerden hun macht als uitkomst van geniaal ondernemerschap en als noodzakelijk bijproduct van technische vooruitgang.[338] Shoshana Zuboff laat zien dat zij zich beriepen op wetten van de economie en de evolutie. Wetten van de overheid waren niet nodig, omdat de wetten van evolutie, kapitaal en vraag en aanbod dat werk reeds deden.[339] Rondom veiligheid op de werkvloer betoogden eigenaren van fabrieken dat deze de verantwoordelijkheid van werknemers zelf was.[340] En de veiligheid van auto's zou niet de verantwoordelijkheid zijn van de producenten maar van gebruikers.

Nieuwe wet- en regelgeving

Wanneer bij een nieuwe technologie de noodzaak van regulering duidelijker wordt, rijst de vraag in hoeverre specifieke wetgeving noodzakelijk is of dat bestaande wetgeving volstaat om ook de nieuwe gevallen te adresseren. Wanneer specifieke regulering of wetgeving nodig werd geacht, bleek het in verschillende gevallen mogelijk om dat zelfs op internationaal niveau succesvol te regelen en de kwalijke effecten en toepassingen van nieuwe technologieën tegen te gaan. Het Genève Protocol van 1925 is daarvan een voorbeeld. Nadat in de Eerste Wereldoorlog chemische wapens waren gebruikt, realiseerde dit protocol een verbod op het gebruik van chemische en biologische wapens.[341] Het *Montreal Protocol on Ozone Depleting Substances* uit 1987 combineerde op succesvolle wijze restricties op het gebruik van bepaalde substanties en stimuleerde tegelijkertijd technologische alternatieven.[342] Illustratief in dit verband zijn ook de afspraken die landen, nu bijna vijftien jaar geleden, binnen de Raad van Europa hebben gemaakt om onlinekinderpornografie tegen te gaan.[343] De vijfde opgave, die we hierna bespreken (paragraaf 3.6), gaat expliciet over de internationale dimensie van maatschappelijke inbedding. In het kader van regulering is het belangrijk om op te merken dat de begrenzing van technologie om gevaren middels wetgeving te mitigeren niet alleen nationaal maar ook internationaal succesvol kan zijn.

---

338 Taplin 2017: 8-9.
339 Zuboff 2019: 106-107.
340 Een grote brand in een kledingfabriek in New York in 1911 leidde tot opschudding en uiteindelijk dwingende regels voor brandveiligheid. Gordon 2016: 271-272.
341 Floridi 2014: 203.
342 Juma 2016: 302.
343 Via zowel het Cybercrime-Verdrag van de Raad van Europa (door Nederland geratificeerd in 2006) als het Verdrag van Lanzarote van de Raad van Europa (door Nederland geratificeerd in 2010) alsmede EU-Richtlijn 2011/93/EU.

Zoals het Collingridge-dilemma al stelt, is het heel lastig, zeker in een vroeg stadium, om duidelijk te krijgen welke soort regels nodig zijn. Sommige regels kunnen de voordelen van een nieuwe technologie namelijk ondermijnen. Een voorbeeld daarvan zijn de Red Flag Acts in het Verenigd Koninkrijk in de tweede helft van de negentiende eeuw. Die schreven onder het mom van verkeersveiligheid voor dat iemand met een rode vlag voor een mechanisch voertuig uit moest lopen.[344] Daarmee werd de snelheid enorm beperkt en de voordelen van het nieuwe vervoersmiddel dus ondermijnd.

Diverse en flexibele instrumenten

Een belangrijke les uit de historie van de regulering van systeemtechnologieën is dat er geen *'silver bullets'* zijn: geen enkele maatregel kan een nieuwe technologie volledig op verantwoorde wijze in de samenleving inbedden. Zoals we bij de auto hebben gezien, gaat het hier om een jarenlang proces waarbij steeds moet worden ingespeeld op nieuwe vraagstukken en gevaren. In 1957 werd in Nederland voor het eerst een maximumsnelheid ingevoerd voor de bebouwde kom. Pas in 1974 volgde die voor de snelweg, toen daar door de toename van het aantal auto's gevaarlijke situaties ontstonden. Gordels werden in 1975 verplicht voor mensen op de voorbank en in 1992 ook voor passagiers achterin. Het keurmerk van de APK werd pas in 1982 verplicht voor alle auto's. Het proces van de auto in en met regels inbedden is bovendien nog steeds niet af. Het is een lerend proces dat na verloop van tijd steeds concretere vorm aanneemt.

De geschiedenis leert niet alleen dat er geen *silver bullets* zijn, maar ook dat er een patroon is in de mate waarin de overheid intervenieert. Veelal is er in eerste instantie een voorkeur om prudent te beginnen met de meest flexibele instrumenten, wordt gaandeweg kennis en ervaring opgedaan en gedurende de tijd opgeschoven naar meer dwingende instrumenten.

Bij flexibele instrumenten kunnen we aan verschillende zaken denken. In de eerste plaats aan wetgevingsanalogieën. Wanneer een nieuw veld ontstaat, zoals gebeurde bij biotechnologie en nanotechnologie, wordt voor analogieën gekeken naar regulering op andere terreinen; in het geval van nanotechnologie was dat bijvoorbeeld de chemische industrie.[345] Andere instrumenten die flexibiliteit bieden en worden ingezet, zijn experimenteerwetten, *soft law* en zogenoemde '*regulatory sandboxes*' waarin nieuwe businessmodellen uitgeprobeerd kunnen worden.

344 Juma 2016: 295.
345 Lee en Vaughan 2010: 193-218.

Het informatieprobleem bij een nieuwe technologie is ook aan te pakken via publiek-private samenwerking. Dat is internationaal een steeds gangbaarder model waar we bij de volgende opgave op terug zullen komen. Het model wordt met name toegepast op hele technische domeinen waarbij expertise uit de private sector van groot belang is. De International Organization for Standardization (ISO) is een voorbeeld van een dergelijke publiek-private samenwerking.[346] Ook in biotechnologie worden allerlei zachtere governance-instrumenten ingezet waarbij onderzoekers, overheden en bedrijven samenwerken en de beste aanpak rondom een nieuwe technologie aftasten.[347]

Toezichthouding

Regulering gaat niet alleen over wetgeving en standaarden, maar ook over het toezicht daarop en de handhaving ervan. Ook daar geldt dat er zeker in de vroege fasen van een nieuwe technologie dynamisch en lerend op ingespeeld moet worden. Een specifiek vraagstuk voor het toezicht van systeemtechnologieën volgt uit het generieke karakter van de technologie. Dat betekent dat het toegepast kan worden in heel verschillende contexten die elk hun eigen regels, waarden, principes en historie hebben. Dat maakt het lastig om in wetten en toezichtsarrangementen alle mogelijke toepassingen centraal te behandelen.

We kijken opnieuw naar het voorbeeld van elektriciteit. Rondom een paar zaken spelen hier universele vraagstukken, zoals het type spanning en de bekabeling. Voor het overgrote deel treedt elektriciteit echter het leven van burgers binnen via heel verschillende producten, van fabrieken, straatverlichting en tandenborstels tot roltrappen en computers. Verreweg de meeste regels omtrent elektriciteit hebben dan ook betrekking op die verschillende toepassingen. Daar komt bij dat generieke technologieën zich vaak kenmerken door hun *dual use*-karakter, waarmee wordt bedoeld dat dezelfde technologie zowel voor militaire als voor civiele doeleinden ingezet kan worden.[348] Beide contexten vragen een heel ander soort regels en handhaving. Specifieke domeinkennis is daarom altijd vereist om systeemtechnologieën ook op het niveau van de toepassing te kunnen reguleren.

Ook de instituties en instanties die de toepasselijke spelregels handhaven, zullen zich moeten verhouden tot de inbedding van een systeemtechnologie. Zeker wanneer de effecten van een nieuwe technologie op de samenleving nog niet zijn uitgekristalliseerd, is de rol van de rechtsspraak als controlerende macht

346 Leung 2019: 17.
347 Leung 2019: 227.
348 Volgens een schatting is in de ruimtevaart bijvoorbeeld 95 procent van alle technologie *dual use* (Leung 2019: 66).

niet te verwaarlozen. Illustratief is een uitspraak uit 1995 van de Amerikaanse rechter inzake cryptografie. De rechter oordeelde dat het in strijd was met het recht op vrije meningsuiting, en daarmee met een vitale component van de democratie, om het verspreiden van encryptiesoftware te verbieden.[349]

Ook het parlement speelt een wezenlijke rol bij het schuren en schaven aan regels en richtlijnen. Zo kunnen parlementsleden vanuit hun controlerende macht misstanden of vraagstukken aankaarten. Als controleur van de uitvoerende macht kan het parlement bovendien technologie politiseren. Neem wederom de geschiedenis van encryptie. Terwijl de Amerikaanse regering en uitvoeringsinstanties als de NSA en de FBI de verspreiding van encryptie zoveel mogelijk wilden tegenhouden, verdedigden leden van het Amerikaanse Congres vaak de rechten van burgers tegenover de staat.[350] Een belangrijke voorwaarde voor rechters en parlementen om deze controlerende functie te vervullen is dat zij voldoende geëquipeerd zijn, met middelen en kennis, om het gebruik van nieuwe technologieën te controleren. In de VS speelde bijvoorbeeld het Office of Technology Assessment van 1972 tot 1995 een belangrijke rol om het Congres hierin bij te staan. Het Rathenau Instituut speelt in ons land een vergelijkbare rol.

Een groeiende rol voor de overheid

Het voorgaande illustreert dat de rol van de overheid, en daarmee van wetgeving, democratische controle en toezicht, groter wordt naarmate een systeemtechnologie meer in de samenleving ingebed raakt, niet in de laatste plaats omdat ook de effecten ervan duidelijker worden. De technologie is dan bovendien dermate ingeburgerd dat de samenleving niet gemakkelijk meer zonder kan. Daarmee wordt de technologie (of aspecten ervan) ook meer en meer als een publiek goed beschouwd, soms zelfs als een utiliteit. Denk aan het openbaar vervoer, het elektriciteitsnetwerk, het wegennet en internetkabels.[351] Ook het eerdergenoemde machtsprobleem speelt een rol bij het gegeven dat de overheid met het verstrijken van de tijd een belangrijkere rol speelt dan in het vroege begin, waarin het vooral private ondernemingen zijn die de technologie vormgeven.

Wat betreft de laatstgenoemde aanleiding voor een grotere rol voor de overheid, het machtsprobleem, zien we dat overheden op verschillende manieren de historische giganten van systeemtechnologieën en daarmee de voorgangers van de

349 Schulz en Van Hoboken 2016.
350 Schulz en Van Hoboken 2016.
351 In de eerste helft van de twintigste eeuw zijn de spoorwegen bijvoorbeeld in landen als Canada, Duitsland, Frankrijk, Nederland, Zweden, Spanje en het Verenigd Koninkrijk genationaliseerd (Van der Vleuten et al. 2017: 74).

huidige Big Tech, hebben aangepakt. Zo werd de macht van de *robber barons* aan de kaak gesteld tijdens het zogenoemde 'Progressieve Tijdperk'. In de VS werd in 1890 de Sherman Act ingevoerd tegen de macht van grote 'trusts'.[352] President Theodore Roosevelt gebruikte die wet om Standard Oil van Rockefeller en Northern Securities van Morgan op te breken.[353]

De overheid kan niet alleen hun macht aanpakken, maar ook voorwaarden stellen aan bedrijven om het publiek belang te dienen. Met de Rural Electrification Administration dwong de Amerikaanse overheid elektriciteitsbedrijven om ook de minder rendabele gebieden van het platteland van hun diensten te voorzien.[354] Toen AT&T een monopolie had in de Amerikaanse telecommunicatie, moest het aan strikte eisen voldoen, zoals het vrijgeven van octrooien.[355]

Ten slotte is het relevant om op te merken dat er significante verschillen bestaan tussen landen in hun traditie met en perspectief op een sturende rol van de overheid. In tegenstelling tot de VS kent Europa bij veel van deze nieuwe technologieën al van meet af aan een publieke inmenging.[356] Duidelijk is in ieder geval dat telkens als een systeemtechnologie wordt ingebed in de samenleving, het publieke belang van die technologieën gedurende de tijd aan relevantie toeneemt, waarmee zich een regulerende rol voor de overheid aftekent.

Kernpunten – Opgave 4: Regulering

- Hoewel het in de vroege fase van een technologie het makkelijkst is om regels op te stellen, bestaat daartoe op dat moment vaak onzekerheid en gebrek aan urgentie. Wanneer die urgentie later wel gevoeld wordt, blijkt het moeilijker om regulering in te voeren en bestaande praktijken te veranderen.
- Aanvankelijk wordt bij nieuwe systeemtechnologieën een beroep gedaan op zelfregulering. De machtsconcentratie van een aantal bedrijven en andere misstanden maken gaandeweg echter dwingender wetgeving noodzakelijk.
- Wetgeving van een nieuwe technologie kent geen *silver bullets*. Het is daarom belangrijk het brede palet van instrumenten te gebruiken. Flexibele instrumenten als experimenteerwetten en *soft law*, maar

352 Freeman en Louçã 2001: 342.
353 Taplin 2017: 115.
354 Gordon 2016: 315.
355 Taplin 2017: 259.
356 Bakker en Korsten 2021.

ook publiek-private samenwerking voor standaarden zijn belangrijk om kennis en expertise op te doen en om te gaan met de genoemde onzekerheden.
- Het generieke karakter van systeemtechnologieën, en daarmee de toepassing in een variëteit aan contexten, vraagt bij toezicht en handhaving vooral om een contextuele benadering.
- De rol en invloed van de overheid bij de inbedding van een technologie verschilt tussen landen maar wordt met de tijd overal dominanter naarmate de technologie een prominentere rol in de samenleving krijgt en het publiek er afhankelijker van wordt.

3.7 Opgave 5: Positionering

De laatste opgave die wij onderscheiden, is positionering; deze betreft het internationale niveau van inbedding. De eerdergenoemde opgaven kennen ook een internationale dimensie. Zo vindt regulering vindt behalve nationaal ook plaats vanuit internationale organisaties. Het engagement van partijen als wetenschappers of activisten heeft eveneens vaak een internationaal aspect. Om twee redenen is internationale positionering ook een afzonderlijke opgave. Allereerst gaat het bij deze opgave om andere spelers dan op het zuiver nationale niveau. Ten tweede spelen in relatie tot het internationale toneel specifieke vraagstukken, zoals het verdienvermogen en de veiligheid van ons land. De centrale vraag bij positionering is dan ook: hoe verhouden wij ons internationaal? (figuur 3.7)

Figuur 3.7 Opgave 5: Positionering

Economische competitiviteit

Een van de karakteristiekste aspecten van systeemtechnologieën in de internationale context is de vaak gehanteerde benadering van de technologie als een race tussen landen. Het idee ontstaat dat landen die vooroplopen in de ontwikkeling en toepassing van de technologie, allerlei voordelen hebben boven andere landen. Die zullen op hun beurt dan weer hun best moeten doen om niet achter te blijven in die race.

Dit frame van een race legt de nadruk op internationale competitie en kan daarmee op gespannen voet komen te staan met normatieve vraagstukken over de technologie. Ten tijde van de Industriële Revolutie werd bijvoorbeeld vanuit het Europese continent met argusogen gekeken naar de economische en technologische ontwikkeling van Engeland. Britse stoommachines maakten veel indruk. Er werd zelfs gesproken van 'het rijk van Vulcanus', de Romeinse god van het vuur, en ook de spoorlijnen, schoorstenen en fabrieken leidden tot vergelijking met de architectuur van het Romeinse Rijk. Dit model was indrukwekkend, en tegelijkertijd ook afstotend. Engeland werd materialisme en hebzucht verweten, wat bijdroeg aan een gevoel van Anglofobie.[357] Met eenzelfde ambiguïteit werd bij latere technologieën gekeken naar Duitsland en de VS. Het idee ontstond zo dat technologisch leiderschap ten koste ging van allerlei fundamentele waarden.

De geschiedenis leert dat een succesvolle ontwikkeling en toepassing van een nieuwe systeemtechnologie bijdraagt aan de competitiviteit van een land, omdat deze als generieke technologie verbeteringen door de gehele economie en samenleving faciliteert. Nationale strategieën met publieke investeringen in infrastructuur en onderwijs kunnen daaraan bijdragen. Zo droeg een gecoördineerde aanpak om wetenschap en industrie met elkaar te verbinden bij aan de economische inhaalslag van Duitsland eind negentiende eeuw. Ook in Oost-Azië hebben in de twintigste eeuw publieke investeringen in nieuwe technologieën bijgedragen aan de economische opkomst van landen als Japan, Zuid-Korea, Taiwan en China.[358]

Militaire verhoudingen

Het competitieve voordeel dat systeemtechnologieën aan landen bieden, is niet alleen economisch. Ook op militair gebied heeft leiderschap in grote nieuwe technologieën bijgedragen aan een sterke positie in de internationale orde. De spoorwegen faciliteerden de Pruisische overwinning op Frankrijk in 1871[359] en

357 Bakker en Korsten 2021: 9
358 Johnson 1982; Wade 2018; Amsde 1989; Zhang 2012.
359 Bousquet 2009.

vervulden een belangrijke rol bij de kolonisatie van de wereld door Europese landen.[360]

Ook in de twintigste eeuw speelden investeringen in de ontwikkeling en toepassing van nieuwe technologieën een centrale rol in conflictsituaties. Tijdens de Tweede Wereldoorlog stonden Britse en Amerikaanse codekrakers als Alan Turing en onderzoekers naar ballistiek, die met hun werk de basis legden voor de computer, tegenover de Duitse raketten van Werner von Braun. Toen in 1957 de Sovjet-Unie de Sputnik 1 lanceerde, was er een wijdverspreide vrees dat de VS de Koude Oorlog zou verliezen omdat het land in de technologische race het onderspit zou delven. Een jaar later leidde de Defense Reorganization Act tot het instellen van ARPA, dat later DARPA genoemd zou worden. Deze onderzoekstak van het Amerikaanse leger zou later verantwoordelijk zijn voor allerlei nieuwe technologieën als GPS en het internet. Ook werd in 1958 NASA door de National Aeronautics and Space Act opgericht, waar onder andere Werner von Braun, die na de oorlog in Operation Paperclip naar Amerika was gehaald, aan Amerikaanse ruimtevaartechnologie zou werken.[361] In de jaren zestig maakte de regering-Kennedy een Amerikaans mondiaal satellietsysteem tot een nationale prioriteit. President Eisenhower steunde eveneens het technologisch leiderschap van Amerikaanse bedrijven omwille van de geopolitieke rivaliteit in de Koude Oorlog.[362]

Pogingen tot nationalisatie

Nationale strategieën voor systeemtechnologieën hebben dus bijgedragen aan het verdienvermogen en de geopolitieke positie van landen. Tegelijkertijd zijn er ook grenzen aan dit beeld van een mondiale race. Het idee dat er een absolute winnaar is, een land dat de technologie weet te monopoliseren en van daaruit een blijvend voordeel krijgt op andere landen, is vanuit historisch oogpunt namelijk ongefundeerd. De ontwikkeling van een systeemtechnologie blijkt doorgaans een internationale aangelegenheid waar verschillende landen aan hebben bijgedragen.

Aan de verbrandingsmotor bijvoorbeeld hebben de Zwitser François Isaac de Rivaz, de Belg Jean Joseph Etienne Lenoir, de Duitsers Nikolaus Otto, Karl Benz, Rudolf Diesel en de Amerikanen George Brayton en George B. Selden allemaal bijgedragen. Ook de ontwikkeling van elektriciteit was een internationale aangelegenheid.[363] Voor de stoommachine gold weliswaar dat die

360 Diogo en Van Laak 2016.
361 Weinberger 2018.
362 Leung 2019: 79-83.
363 Bakker en Korsten 2021: 16.

goeddeels in Engeland is ontwikkeld, maar het land wist dit niet in een blijvend voordeel te vertalen. Ondanks een aanvankelijke achterstand bleek de motor van Amerikaanse origine, Corliss, in de loop der tijd superieur en wist deze ook de Engelse markt te veroveren.[364]

Rondom systeemtechnologieën bestaat bovendien een internationale oriëntatie die sterk wordt gedreven door wetenschappelijke onderzoekers. Belangrijk daarbij is het belang dat zij hechten aan open toegang tot kennis en bijdragen aan internationale conferenties en vakbladen. De pogingen om systeemtechnologieën te 'nationaliseren' worden daarom veelal gedreven door overheden en niet door de wetenschap.

Zo legden in het geval van elektriciteit de Britten, in reactie op de opkomst van de VS en Duitsland, met behulp van de Italiaanse ingenieur Marconi, draadloze telegrafische netwerken aan om de internationale communicatie te domineren. Met een 'imperiale keten' hoopten zij deze te beheersen. Later poogde de VS hiervoor een alternatief op te zetten en blokkeerde de Amerikaanse marine de verkoop van geavanceerde technologie van GE. De Radio Corporation of America was opgericht om wereldwijd radiohegemonie te verkrijgen. Zowel de Britse als de Amerikaanse pogingen faalden en landen als Frankrijk en Duitsland zetten hun eigen radiostations op voor nationale communicatie.[365]

Het verleden leert niet alleen dat pogingen tot nationalisering van een nieuwe technologie keer op keer mislukken, maar ook dat die vaak zelfs averechts werken. Door de politisering van de technologie worden andere landen namelijk gemotiveerd om hun eigen alternatieven op te zetten. Politisering is bovendien slecht voor de marktpositie van het land dat dit beleid voert, omdat internationale klanten huiverig worden van de politieke inmenging of omdat de meest geavanceerde producten niet meer verkocht mogen worden. Als gevolg daarvan kan dit beleid de marktpositie van het leidende land juist ondermijnen.

Een voorbeeld hiervan komt uit de ruimtevaartindustrie en betreft competitie tussen de VS en China, en is daarmee relevant voor de hedendaagse concurrentie tussen beide landen op het gebied van AI. Uit angst voor Chinese spionage onthield het Amerikaanse Congres in 1989 goedkeuring voor exportlicenties voor Amerikaanse satellieten aan boord van Chinese raketten. Een rapport uit 1998 concludeerde dat Chinese acquisities van technologie de Amerikaanse veiligheid in gevaar hadden gebracht en dat exportcontroles op satellieten versterkt moesten worden. In 1999 volgde de strenge Strom Thurmond Act.

364 Bakker en Korsten 2021: 27.
365 Bakker en Korsten 2021: 15.

Dit beleid had een negatief effect op de concurrentiepositie van de Amerikaanse satellietindustrie. Terwijl deze in 1995 nog 90 procent van de markt voor satellietcomponenten in handen had, daalde dat aandeel tot 56 procent in 1999. Buitenlandse bedrijven als het Duitse DaimlerChrysler Aerospace en Telesat Canada verbraken de banden met Amerikaanse bedrijven om hun onbetrouwbaarheid en zetten alternatieven op.[366]

Een tweede interessant voorbeeld van mislukte nationalisering komt uit de cryptografie. Ook op deze technologie poogde de Amerikaanse federale overheid grip op te krijgen. De NSA stelde bijvoorbeeld in de jaren tachtig voor dat de gehele Amerikaanse industrie gebruik zou gaan maken van door de NSA ontwikkelde algoritmes ingebed in speciale versleutelde chips. Het vermoeden was echter dat deze niet voor veiligheid bedoeld waren, maar om de NSA toegang tot alle communicatie te verlenen. In 1993 kwam de regering-Clinton met het Clipper-initiatief waarbij bedrijven encryptie moesten gebruiken waarvan de sleutels bij de overheid zouden komen te liggen. Dat leidde tot grote kritiek. Bedrijven klaagden dat zij nooit producten aan het buitenland zouden kunnen verkopen die achterdeuren voor de Amerikaanse veiligheidsdiensten bevatten. Burgerrechtengroepen bekritiseerden de surveillance die dit initiatief mogelijk maakte en onderzoekers toonden aan dat dit systeem helemaal niet technisch robuust was. Later kwam de regering met een hernieuwd Clipper II-initiatief, maar uiteindelijk faalden beide initiatieven.

Een ander instrument dat de Amerikaanse overheid gebruikte om encryptie te domineren, waren exportcontroles. Producten met hele sterke encryptie moesten op basis van een wet uit 1976 goedkeuring krijgen om geëxporteerd te worden. Die goedkeuring werd echter slechts zelden verkregen. Sterke encryptie mocht binnen de VS worden toegepast, maar zwakkere varianten moesten geëxporteerd worden. Als gevolg hiervan verloren Amerikaanse bedrijven positie op de internationale markt. Met een meer en meer globaliserende markt en de beschikbaarheid van *open source*-kennis draaide de Amerikaanse overheid rond de eeuwwisseling uiteindelijk dit beleid van exportcontroles terug.[367]

Wetenschappers en andere onderzoekers die hun kennis in weerzin van de overheid *open source* deelden – een strijd die de 'Crypto Wars' wordt genoemd –, vormden een belangrijke tegenkracht tegen de pogingen van de Amerikaanse overheid om cryptografie te nationaliseren. Dat gold ook voor bedrijven. Terwijl de private sector met de overheid gelieerd kan zijn, geven bovengenoemde voorbeelden aan dat ondernemingen de overheid ook kunnen

366 Leung 2019: 94-99.
367 Leung 2019: 195-199, 217.

tegenwerken als ze hun positie op de internationale markt proberen te beschermen. Na de terroristische aanslag in San Bernardino in 2015 weigerde Apple bijvoorbeeld in te gaan op het verzoek van de FBI om de encryptie van de iPhone van de aanslagplegers op te heffen. Een rechtszaak volgde. Apple beargumenteerde daar dat het verzoek van de Amerikaanse overheid de privacy van alle gebruikers in gevaar zou brengen. Als gevolg van deze zaak ging een reeks van Amerikaanse technologiebedrijven als Whatsapp, Yahoo en Google over op sterke encryptie, wat de FBI-directeur James Comey het *'Going Dark'*-probleem noemde.[368]

Het belang van internationale samenwerking

Tegenover de poging om een systeemtechnologie te nationaliseren, staan inspanningen voor open internationale samenwerking. In allerlei formele en informele verbanden worden vanuit deze oriëntatie standaarden, richtlijnen, codes en principes van goed gebruik opgesteld rondom systeemtechnologieën. Een mooi historisch voorbeeld van het vaststellen van internationale technische standaarden was de bijeenkomst van de British Association in 1861 waarbij afspraken zijn gemaakt voor de tot op heden gebruikte standaarden joule, ohm en ampère.[369] Illustratief is ook de regeling van domeinnamen, die cruciaal was voor de mondiale standaardisatie van de adressering van computers. Het zwaartepunt van de belangstelling hiervoor lag bij universiteiten en in eerste instantie niet bij bedrijven en overheden.[370]

Op het gebied van biotechnologie zetten onderzoekers ook allerlei vormen van zelfregulering op. In relatie tot biotechnologie wijst Colin Scott er bovendien op dat het meerwaarde heeft om bij het opstellen van internationale standaarden, richtlijnen en andere vormen van zelfregulering informele verbanden te benutten. Een te dominante rol van overheden bij het opstellen van standaarden miskent zowel de expertise die elders aanwezig is als de kans om een gevoel van 'eigenaarschap' van regulering te realiseren.[371] Wetenschappers Wolfram Kaiser en Johan Schot hebben laten zien dat ver voor het ontstaan van de Europese Unie de technocratische benadering van experts en industrieassociaties al vanaf de negentiende eeuw een kracht voor verdere Europese eenwording zou blijken.[372]

368 Leung 2019: 208-209.
369 Bakker en Korsten 2021: 12.
370 Zie Olsthoorn 2015 voor een illustratief overzicht van de eerste ontwikkelingen en pioniers in Nederland.
371 Scott 2007: 19-38.
372 Kaiser en Schot 2014: 294-296.

Toch zijn ook nationale staten erin geslaagd om internationale verdragen voor het gebruik van nieuwe technologieën op te stellen. Zo werd in 1975 de *Biological Weapons Convention* van de Verenigde Naties van kracht, de eerste internationale poging om een hele klasse van wapens te verbieden.[373] Op het gebied van ruimtevaarttechnologie werden vanaf 1967 vijf internationale verdragen getekend over zaken als de vreedzame exploratie van de ruimte, schade door objecten in de ruimte en het militariseren van de maan.[374]

Tegenover de focus op nationale economische en militaire macht staan dus initiatieven voor internationale samenwerking. Belangrijk om op te merken is dat internationale samenwerking op nieuwe technologieën ook vaak het expliciete doel had om vrede te stimuleren. Met dat doel stelde de Italiaan Piero Puricelli in 1921 een Europees snelwegennetwerk voor. Dat doel speelde ook een rol bij de daadwerkelijke realisering hiervan na de Tweede Wereldoorlog.[375] Ook de oprichting van CERN na de Tweede Wereldoorlog diende ertoe om welvaart en samenwerking te promoten en onderzoekers met niet-militair onderzoek te faciliteren.[376]

Tegelijkertijd kon de nadruk op internationale samenwerking soms ook dienen om competitie te verhullen. De samenwerking tussen de VS en de Sovjet-Unie om "met gezamenlijke moeite het universum te meesteren" verhinderde beide om individueel een leidende rol te gaan spelen.[377]

Er is nog een laatste kanttekening te plaatsen bij het idee dat landen rondom een systeemtechnologie in een race met elkaar verwikkeld zijn. Een race suggereert namelijk dat iedereen dezelfde wedstrijd speelt. Uit de geschiedenis van systeemtechnologieën blijkt echter dat landen technologieën ook heel verschillend in kunnen zetten. De ontwikkeling van het elektriciteitsnetwerk was in de VS bijvoorbeeld gedreven door de commerciële belangen van de private sector. In Europa daarentegen werden die netwerken al vroeg gezien als een publieke dienst. Europese huishoudens werden hierdoor sneller en tegen lagere kosten aan het net verbonden dan hun Amerikaanse tegenhangers, terwijl dat land vooropliep in commerciële toepassingen. Het doel en de aard van toepassing van een nieuwe technologie staat dus niet bij voorbaat vast.[378]

373 Kaiser en Schot 2014: 134.
374 Kaiser en Schot 2014: 82.
375 Bakker en Korsten 2021: 17-18.
376 Van der Vleuten et al. 2017: 96-97.
377 Leung 2019: 87-90.
378 Bakker en Korsten 2021: 12.

Kernpunten – Opgave 5: Positionering

- Bij nieuwe systeemtechnologieën ontstaat veelal het beeld van een wereldwijde race. Stimulering van een nieuwe technologie door middel van strategische programma's draagt bij aan het verdienvermogen van landen en aan hun strategische positie. Omwille van economische en geopolitieke motieven is het dus van groot belang om te investeren in nieuwe systeemtechnologieën.
- Tegelijkertijd is het beeld van een race misleidend. Juist bij een systeemtechnologie zijn de ontwikkeling en de vooruitgang altijd een internationale aangelegenheid. Pogingen om de ontwikkeling te nationaliseren en andere landen af te sluiten hebben meestal niet gewerkt en zijn zelfs contraproductief gebleken.
- Internationale samenwerking en ontwikkeling van universele standaarden draagt bij aan de inbedding van een nieuwe systeemtechnologie.
- Het beeld van een race verhult ten slotte de diversiteit tussen landen in de omgang met een systeemtechnologie en de verschillende waarden die ten grondslag kunnen liggen aan de vormgeving en inzet van die technologie.

We hebben in dit hoofdstuk vijf opgaven besproken die historisch cruciaal zijn gebleken voor de maatschappelijke inbedding van systeemtechnologieën. In het volgende deel van dit rapport werken we deze opgaven uit voor AI, en bekijken we per opgave de huidige dynamieken en de betekenis daarvan voor de maatschappelijke inbedding van deze nieuwe systeemtechnologie.

Deel 2

Vijf opgaven: bespreking van de opgaven voor de maatschappelijke inbedding van AI

4. Demystificatie

De eerste maatschappelijke opgave die we bespreken is demystificatie. Deze opgave gaat over de beeldvorming rondom een nieuwe technologie bij het bredere publiek. Juist omdat systeemtechnologieën zo breed ingezet kunnen worden en door hun generieke karakter een zekere ontastbare kwaliteit krijgen, spreken ze bijzonder tot de verbeelding. In hoofdstuk 3 bespraken we dat daarmee zowel te hooggespannen verwachtingen als overtrokken angstbeelden kunnen ontstaan. Beide kunnen de inbedding van een technologie in de maatschappij in de weg staan. Demystificatie, als het tegengaan van overspannen voorstellingen van AI, is van belang om reële kansen én risico's niet uit het oog te verliezen en draagt zo bij aan de kwaliteit van het debat over AI. Van beelden die de aandacht trekken, naar vragen die de aandacht verdienen.

In het vorige hoofdstuk zagen we al kort hoe de introductie van een nieuwe systeemtechnologie als elektriciteit gepaard kan gaan met mythevorming. Ook bij de opkomst van AI zien we een dergelijke dynamiek. We lichten een aantal prominente AI-mythen uit: voorbeelden van een te optimistische, pessimistische of simpelweg gebrekkige voorstelling van wat AI is. We laten zien waar de misvattingen liggen en waar het werkelijke vraagstuk. Daarmee demystificeren we enkele onrealistische en ongenuanceerde beelden over AI. Tot slot bekijken we hoe deze opgave er op maatschappelijk niveau uitziet. Hoe kunnen we er als samenleving voor zorgen dat onze omgang met AI niet wordt geleid door onrealistische beelden? Oftewel: *Waar hebben we het over?*

4.1 Mythevorming rondom AI

Utopie en dystopie

In de geschiedenis van systeemtechnologieën is een aantal patronen te herkennen als het gaat om de publieke beeldvorming. Een eerste patroon is het ontstaan van utopische voorstellingen enerzijds en doemscenario's anderzijds. Ook rondom AI zijn die twee extremen zichtbaar. "*We're at the beginning of a golden age of AI*," aldus Jeff Bezos, CEO van Amazon. Elon Musk ziet dat anders: "*With AI we are summoning the demon*."[379] Hun uitspraken zijn illustratief voor twee uiterste sentimenten die gepaard gaan met de opkomst van AI. Sommigen onthalen de technologie als de ultieme technologische verlossing, anderen als een grote bedreiging voor de mens. Volgens Rodney Brooks, robotpionier, berusten veel utopieën en angstbeelden op misvattingen over wat AI is: "[…]

379 Musk gelooft in een toekomst die veel weg heeft van *The Matrix*. In een interview werd hem gevraagd wat hij in de toekomst zou vragen aan een systeem met *artificial general intelligence*. Zijn vraag: "*What is outside the simulation*?" (Fridman, 16 augustus 2019).

having ideas is easy. Turning them into reality is hard. Turning them into being deployed at scale is even harder."[380] Volgens Brooks liggen mythen over AI vaak ten grondslag aan onrealistische verwachtingen over wat ze ons ten goede of ten kwade zal brengen.

Groot vertrouwen in de heilzame werking van AI kan zich manifesteren als *technosolutionisme*. Evgeny Morozov beschrijft daarmee de neiging om complexe sociale fenomenen om te dopen tot een vraagstuk waar technologie het antwoord op is. Het oplossen van problemen wordt dan een kwestie van het juiste algoritme erop loslaten.[381] Deze '*silicon mentality*', zoals Morozov deze neiging eerder karakteriseerde, zien we bij uitstek terug rondom AI. Astro Teller, hoofd van Alphabet's technologielab 'X', stelde dat er 90 procent kans is dat 'slimme' machines specifieke problemen in de samenleving kunnen oplossen.[382] De oprichter van DeepMind Demis Hassabis voorspelt dat bovenmenselijke intelligentie grote problemen, van klimaatverandering tot ongeneeslijke ziekten, zal oplossen.[383]

Het andere uiterste speelt rondom de opkomst van AI: een diep wantrouwen jegens alles wat met algoritmes en automatisering te maken heeft. Ideeën rondom ontmenselijking, massawerkeloosheid of zelfs existentiële bedreiging van de mens liggen ten grondslag aan grote zorgen over AI. Net zoals we bij elektrische straatverlichting zagen, wordt ook AI in verband gebracht met de angst voor een Big Brother-achtige samenleving waarin digitale technologie wordt gebruikt om ons continu in de gaten te houden. AI wordt ook opgenomen in bestaande complottheorieën, bijvoorbeeld in het kader van 5G en de zorgen rondom straling en privacy daaromtrent.[384] In het voorjaar van 2020 ging er zelfs een verhaal rond dat COVID-19-vaccins ons DNA zouden manipuleren en ons zouden aansluiten op een AI-systeem dat continu informatie over ons ontvangt.[385]

Uit een wereldwijde survey in opdracht van het World Economic Forum blijkt dat vier op de tien mensen zich zorgen maken over AI.[386] Onderzoek naar de houding van Amerikanen ten aanzien van de technologie wijst uit dat de

380 Brooks, 1 januari 2018.
381 Morozov 2013.
382 Tilley, 24 maart 2016.
383 Marcus en Davis 2019.
384 Martin L. Pall, emeritus hoogleraar biochemie, verbindt zijn waarschuwing ten aanzien van 5G-straling aan zorgen over artificiële intelligentie (Pall 2019). Zie verder: Andersen, september 2020 en Halpern, 26 april 2019.
385 Reuters, 24 april 2020
386 Ipsos 2019.

meeste Amerikaanse burgers de verdere ontwikkeling van AI steunen, maar een uiteindelijk negatieve impact verwachten naarmate ze 'intelligenter' wordt.[387] Nederlanders associëren AI in de eerste plaats met 'computers' en 'robots'. Uit een enquête blijkt dat meer dan de helft van de burgers zowel positieve als negatieve gevoelens bij AI heeft. Ze zien grote kansen op het gebied van zorg en het verbeteren van veiligheid, maar vrezen ook voor ongewenste gevolgen van AI. Vooral bij lager opgeleide Nederlanders leeft een zekere angst voor het verlies van banen en de menselijke factor. Hoger opgeleiden maken zich met name zorgen over een gebrek aan controle op AI-systemen en de schending van privacy.[388]

Publieke evenementen

Een ander historisch patroon dat samenhangt met het ontstaan van vertekende beelden rondom generieke technologieën als AI, is de invloed van evenementen op de publieke perceptie. In reactie op angstbeelden rondom opkomende systeemtechnologieën werden livedemonstraties georganiseerd om te laten zien dat een technologie juist betrouwbaar en tot spectaculaire dingen in staat was. In het vorige hoofdstuk kwamen al historische voorbeelden voorbij van publieke wedstrijden en tentoonstellingen waarin toepassingen van een nieuwe technologie bij het grote publiek werden geïntroduceerd, zoals de demonstratie van elektriciteit.

Ook rondom AI komt dit patroon van wedstrijden en publieke demonstraties terug. Veel van de mijlpalen in de ontwikkeling van AI waren zelfs een combinatie van die twee: het moment dat IBM's schaakcomputer Deep Blue won van wereldkampioen Garry Kasparov, de winst van IBM's Watson in *Jeopardy!*, de zege van AlphaGo op twee Go-wereldkampioenen, DeepMind's Agent57 die mensen in 57 Atari-videogames kan verslaan – allemaal wedstrijden waarin wordt gedemonstreerd waartoe AI in staat is. Demonstraties die des te krachtiger zijn omdat de 'intelligentie' van AI hier wordt afgezet tegen die van menselijke kampioenen. Zelfs wanneer AI het aflegt tegen een menselijke tegenstander, kan de krachtmeting indrukwekkend zijn: in 2019 nam IBM's Watson het op tegen 's werelds beste debater. Watson verloor, maar de vertoning kan desondanks als groot succes worden gezien. Alleen al het feit dat een computer de mens kan uitdagen in zoiets complex als een debatwedstrijd was genoeg om aan het publiek te laten zien hoe ver de AI-technologie al is en reuring te creëren over de toekomst ervan.

387 Zhang en Dafoe 2019.
388 Schothorst en Verhue 2018.

Er worden ook wedstrijden tussen AI-systemen georganiseerd. DARPA heeft in het verleden de DARPA Grand Challenge georganiseerd, een competitie voor zelfrijdende auto's. Tussen 2012 en 2015 kwam daar de DARPA Robotics Challenge bij. Beide competities leverden spectaculaire beelden op van racende autonome voertuigen en robots die fysieke tests uitvoeren. Sinds 1990 wordt ook jaarlijks de Loebnerprijs uitgereikt aan de chatbot die het dichtst in de buurt komt van het slagen in de uitgebreide Turingtest en dus het moeilijkst te onderscheiden is van een mens. De gouden en zilveren medaille zijn echter nog nooit gewonnen; alleen de bronzen medaille voor 'minst teleurstellende' bot is tot nu toe uitgereikt.[389]

De kracht van livedemonstraties komt ook terug bij AI. Veel conferenties openen bijvoorbeeld met een 'gesprek' met een robot. De presentator stelt de robot op het podium enkele vragen waarop hij vaak een bijdehand antwoord van de robot terugkrijgt. Daarmee wordt gesuggereerd dat de robot een persoonlijkheid heeft en als de robot faalt, wordt dat vaak eerder als iets menselijks dan als iets technisch afgedaan. Tijdens een presentatie van AI-robot CLOi van elektronicabedrijf LG bleef een antwoord van de robot pijnlijk genoeg tot drie keer toe uit op die vragen die de presentator stelde. Zijn verklaring aan het publiek was dat 'zelfs robots weleens een slechte dag hebben' of dat ze hem niet aardig vond en 'blijkbaar niet met hem wilde praten'. Ook Apple en Google introduceerden hun stemgestuurde assistenten tijdens livedemonstraties. BostonDynamics publiceert indrukwekkende filmpjes om aan het grote publiek de souplesse van hun robots te demonstreren. In een van de laatste filmpjes danst de hele BostonDynamics op een toepasselijk nummer van *The Contours:* "*Do you love me – now that I can dance?"*

Dergelijke demo's spreken letterlijk tot de verbeelding: het publiek wordt niet langer gouden bergen beloofd, maar krijgt ze te zien. Tegelijkertijd kan dit soort evenementen het publiek vaak op het verkeerde been zetten als het gaat om de werkelijke stand van de technologie. De video van BostonDynamics is voor zover bekend niet gemonteerd. Het zijn dus echt de robots die deze dansbewegingen maken. Maar echt dansen is het natuurlijk niet: elke beweging is minutieus voorgeprogrammeerd.[390] De suggestie dat de robots hier de mens evenaren in dans, is in die zin misleidend. Volgens de eerdergenoemde Rodney Brooks leidt dit soort demonstraties tot misvattingen over AI.[391] Het publiek ziet

389 Luciano Floridi, een van de juryleden in 2008, stelde de chatbot de vraag: "als we elkaars hand vasthouden, wiens hand heb ik dan vast?", waarop de computer het onzinnige antwoord gaf: "We leven in de eeuwigheid. Dus, ja, nee. We geloven niet" (Floridi et al. 2009).

390 Ackerman, 7 januari 2021.

391 Ford 2018. Zie ook: Association for Advancing Automation, 25 januari 2018.

alleen wat er op het podium gebeurt en niet wat er achter de schermen allemaal aan mensenwerk schuilgaat om de computer te laten doen wat hij doet.

In de inleiding van dit rapport verwezen we naar het artikel uit *The Guardian* dat in 2020 veel aandacht trok, met de titel: *A robot wrote this entire article. Are you scared yet, human?*[392] Het hele artikel was gegenereerd door nieuwe taalverwerkingssoftware genaamd 'GPT-3', *generative pre-trained transformer 3*. GPT-3 kan met relatief weinig input natuurlijke teksten produceren en het artikel uit *The Guardian* werd opgevoerd als bewijs daarvan. In een tekst die niet te onderscheiden is van menselijk werk wordt geprobeerd om de lezer ervan te overtuigen dat hij niet bang hoeft te zijn voor robots en AI: "*I am here to convince you not to worry. Artificial intelligence will not destroy humans. Believe me.*" Velen waren hiervan onder de indruk en voelden zich getuigen van een voorschot op de toekomst. Achteraf bleek echter dat menselijke redacteuren een cruciale rol hadden gespeeld bij de totstandkoming van het artikel. Met GPT-3 waren in totaal acht essays geproduceerd, waaruit menselijke redacteuren delen hebben geselecteerd en daarmee het uiteindelijke artikel hebben gecomponeerd.[393] Iemand vergeleek het met "het selecteren van zinnen uit spam-berichten, die samenvoegen en claimen dat de spammers Hamlet hebben geschreven."[394]

De prestaties van een AI-systeem worden in demonstraties dus dikwijls overdreven. Wat spontaan overkomt, is vaak voorgeprogrammeerd of anderszins voorbereid door mensen. De rol die mensen spelen, blijft vaak echter letterlijk en figuurlijk buiten beeld. Bovendien vinden de opvoeringen doorgaans plaats in een extreem gecontroleerde setting. Wat het publiek te zien krijgt, is dus meestal misleidend en zegt weinig over hoe het systeem functioneert in de ongecontroleerde en sterk veranderlijke omgevingen van alledag. Door buiten beschouwing te laten wat ervoor nodig is om AI te laten doen wat het op het podium doet, worden mensen verleid om dat wat ze zien te generaliseren naar robuuste en breed inzetbare vaardigheden van AI-systemen. Op die manier kunnen publieke demo's of 'bewijsstukken' van AI in actie bijdragen aan onrealistische voorstellingen van wat AI op dit moment of in de nabije toekomst kan.

De kracht van woorden

Een laatste patroon in het kader van mythevorming rondom systeemtechnologieën heeft te maken met het gebruik van bepaalde woorden. Eerder haalden we het voorbeeld van elektrocutie aan, waardoor elektriciteit werd geassocieerd met dodelijk gevaar. Rondom AI worden inmiddels ook termen gebruikt die een

392 GPT-3, 8 september 2020.
393 GPT-3, 8 september 2020.
394 Geciteerd in: Macaulay, 8 september 2020.

sterk associatief karakter hebben en daarmee direct een bepaald beeld oproepen. Het simpelste voorbeeld hiervan is het gebruik van de term 'intelligentie', waarmee dat wat AI kan wordt verbonden aan onze eigen vermogens. Die associatie kan misvattingen en daardoor ook onjuist gebruik in de hand werken. Hetzelfde geldt voor het gebruik van 'menselijke' werkwoorden als 'denken', 'leren' (*machine learning*), 'redeneren' (*automated reasoning*) en 'observeren' voor wat AI-systemen doen.

Denk hierbij ook aan het geven van menselijke namen of titels aan AI-systemen, zoals de 'robotrechter', 'robotagent' of 'robotdokter'. In lijn daarmee worden AI-systemen soms aangeduid als 'digitale collega'. Daarmee raakt niet alleen onderbelicht dat AI-systemen anders werken dan mensen, maar ook dat het werken *met* een AI-systeem andere processen en vaardigheden veronderstelt dan het werken met menselijke collega's. Het vermenselijken van AI door het gebruik van bepaalde woorden kan dus bijdragen aan vertekende beelden over wat AI nu eigenlijk is. Ook de term 'autopilot' wekt de suggestie van een geautomatiseerde bestuurder, terwijl deze systemen in feite slechts een ondersteunende functie hebben. De aanduiding daarvan als autopilot kan daarom een verkeerd beeld oproepen van wat het systeem doet, zo vindt ook de RDW.[395] Dit heeft als risico dat de noodzaak tot verantwoording hierdoor sneller bij het systeem zélf gelegd wordt, in plaats van bij de afzenders (gebruikers en ontwerpers) van het systeem. Een voorbeeld hiervan is de aanname dat volgers van bepaalde twitteraccounts geautomatiseerd advertenties te zien krijgen, terwijl blijkt dat daar in sommige gevallen zeer bewust gerichte menselijke handelingen aan vooraf kunnen afgaan.[396]

Maar er worden ook termen gebruikt die minder subtiel bepaalde associaties opwerpen. De 'killer robot' of 'killer drone' is daarvan een lichtend voorbeeld. Daarmee wordt de automatisering van wapensystemen geframed als de creatie van moordmachines. Zeker in dit geval van autonome wapensystemen geeft dat de maatschappelijke discussie een sterke richting mee. Een andere beladen term die vaak in de context van AI wordt gebruikt, is 'dataïsme'. De term is populair geworden door Yuval Noah Harrari, die ermee in zijn boek *Homo Deus* verwijst naar een haast religieus geloof in het heil van data en algoritmes.[397] De term is in zwang geraakt en duikt vaak op in het publieke debat om het gebruik van data en AI te framen als een laakbare ideologie waarbij dat wat ons menselijk maakt uit het oog raakt. De frase '*computer says no*', bekend door een sketch van het satirische programma *Little Britain*, wordt eveneens vaak gebruikt om

395 Trouw, 17 oktober 2016.
396 Sheikh 2021.
397 Harrari 2017.

het schrikbeeld op te roepen van een door computers gedomineerd bestel dat flexibiliteit en de menselijke maat ontbeert.

Ook de veelgebruikte term 'black box' is noemenswaardig. Door AI een black box te noemen wordt gesuggereerd dat de mens in het duister tast als het gaat om de werking van zo'n systeem. Het is daarom opmerkelijk dat het Systeem Risico Indicatie (SyRI) van de overheid in eerste instantie de naam 'Black Box' kreeg, waarmee het beeld wordt gewekt dat dit een systeem is waarin geen betekenisvol inzicht mogelijk is.[398] Het beeld van AI als iets ondoorgrondelijks nuanceren we in de volgende paragraaf.

Een andere veel gehoorde frame is het narratief van de 'race' om AI die gewonnen moet worden of die (bijna) verloren is. Virginia Dignum – mede-oprichter van ALLAI stelt dat media en beleidsmakers zijn geobsedeerd door de zogenoemde AI-race. De angst ligt erin dat China deze vermeende race zou kunnen winnen waardoor andere landen zich moeten haasten om niet achter te blijven. Volgens Dignum is dit 'race'-narratief zowel verkeerd als gevaarlijk omdat het de focus legt op competitie en een sfeer van somberheid en wanhoop met zich brengt.[399] Het appèl dat hiermee wordt gedaan op emoties (verliezen), leidt er in ieder geval toe dat overheden wereldwijd enorme bedragen stoppen in innovatie om de race maar niet te verliezen of achterop te raken. Op het frame van een race gaan we in hoofdstuk 8 uitgebreid in.

Het gebruik van specifieke termen en frames kan sterk bepalen hoe mensen over AI denken en spreken, bepalender vaak dan argumenten en feiten. Dat betekent dat beelden zich niet altijd rationeel laten weerleggen. De kracht van woorden moet dus niet worden onderschat. Naast het gebruik van geladen termen hebben we in deze paragraaf ook andere historische patronen geïdentificeerd in de beeldvorming rondom AI. Het imponeren van het publiek door wedstrijden of livedemonstraties, de associatie met een cluster van andere zorgen en overtrokken verwachtingen ten aanzien van wat een nieuwe generieke technologie als AI ons gaat brengen. Te midden van deze cocktail ontstaan vertekende en soms ronduit onrealistische beelden over waar we het nu precies over hebben met AI. Om daar meer helderheid over te krijgen, bespreken we in de volgende paragraaf een paar van de meest voorkomende AI-mythen en laten we zien hoe die ons op het verkeerde been zetten.

---

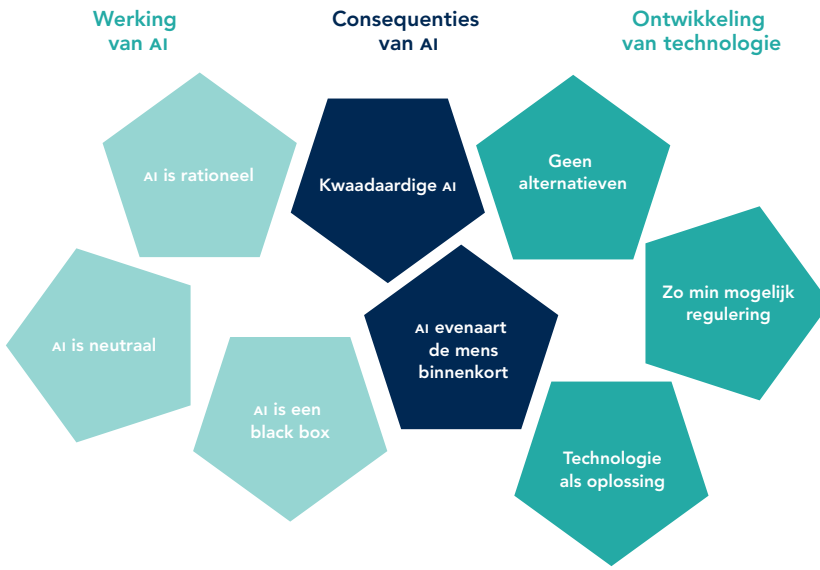
398 In de Staatscourant wordt black box (voorganger van SyRI) gedefinieerd als: "een professionele en beveiligde, organisatorische voorziening waarin door middel van speciale software persoonsgegevens geanonimiseerd worden gekoppeld." Staatscourant 2009, 11, 19 januari 2009.

399 Dignum z.d.

4.2 Hedendaagse mythen over AI

Net als bij eerdere systeemtechnologieën zijn over AI verschillende mythen ontstaan. In deze paragraaf onderzoeken wij de voornaamste daarvan. Een aantal is specifiek op AI gericht en een aantal is van algemenere aard. Eerst kijken wij naar de AI-specifieke beelden, naar de werking en de consequenties ervan. Daarnaast is er nog een categorie van mythen die generieker is en die gaat over de aard van bredere digitale technologie en de manier waarop technologieën als AI ontwikkeld worden door Silicon Valley. Zie figuur 4.1 voor een overzicht.

Figuur 4.1 Overzicht van beelden en mythen over AI



Mythen over de werking van AI

'Kunstmatige intelligentie is neutraal'

Dit is een zeer gangbaar beeld over AI. De gedachte is namelijk dat, in tegenstelling tot mensen, AI-systemen geen last hebben van allerlei zwakten, angsten of vooringenomenheden. In deze context wordt weleens een Israëlisch onderzoek aangehaald dat zou aantonen dat rechters anders oordelen wanneer ze honger hebben dan wanneer ze dat niet hebben.[400] AI heeft nooit honger. Ze is ook nooit moe of met het verkeerde been uit bed gestapt. Van automatische

400 Danziger et al. 2011.

wapensystemen wordt gezegd dat ze nooit emotioneel worden, geen haat voelen en dus ook niet vatbaar zijn voor *overkill*.[401] Ook zou AI neutraal zijn, omdat ze geen last heeft van vooroordelen. Het Amerikaanse systeem COMPAS was bedoeld om het risico op recidivisme van mensen te bepalen. Op de factsheet van het bedrijf staat: "*Objective, standardised instruments, rather than subjective judgments alone, are the most effective methods for determining the programming needs that should be targeted for each offender*".[402]

Zonder emoties, vooroordelen en ideologische overtuigingen zou AI objectievere oordelen kunnen vellen dan mensen. Een hieraan verwant idee is dat de werking van AI apolitiek is. In plaats van ideologisch te strijden over wat gedaan moet worden, kan een AI-systeem een situatie wiskundig optimaliseren waardoor iedereen zonder aanziens des persoons neutraal behandeld wordt.

Het COMPAS-systeem bleek echter de kans op recidive bij zwarte mensen te overschatten en bij witte mensen te onderschatten.[403] AI kent zelf weliswaar geen emoties, vooroordelen of belangen, maar is daarmee nog niet neutraal. De *werking* van AI kan namelijk wel degelijk bevooroordeeld of ideologisch zijn. Ten eerste kunnen er vooroordelen schuilgaan in de gebruikte data – het bekende fenomeen van '*garbage in, garbage out*'. Om algoritmes te trainen, zijn er trainingsdata nodig. Als die data van slechte kwaliteit zijn, bijvoorbeeld omdat ze bevuild zijn, onvolledig of bevooroordeeld, dan zal dat doorwerken in het functioneren van het algoritme. Neem de werking van Google's *search algorithm*. Marcus en Davis noemen verschillende voorbeelden van *bias* die het gevolg zijn van training met bestaande data op het internet. Ze halen een onderzoek uit 2013 aan dat liet zien dat iemand die een typische naam voor een zwart persoon zoals Jermaine googelt, een veel grotere kans heeft om bij advertenties informatie over arrestaties te krijgen dan iemand met een typisch witte naam.

In 2015 labelde Google Photos een aantal Afro-Amerikaanse mensen als gorilla's. Volgens een ander onderzoek geeft de zoekopdracht 'professionele haarstijl voor werk' afbeeldingen van witte vrouwen, terwijl bij 'onprofessioneel' zwarte vrouwen verschijnen. De zoekopdracht 'moeder' levert verreweg de meeste afbeeldingen van witte vrouwen op en bij 'professor' is maar ongeveer 10 procent vrouw.[404] Een HR-algoritme van Amazon bleek systematisch vrouwen uit te sluiten voor banen.[405] Ruha Benjamin noemt een onderzoek uit 2016 waarbij de zoekopdracht 'drie zwarte jongeren' foto's opleverde van arrestaties, 'drie

401 NAVO, 12 december 2019.
402 Broussard 2019: 155.
403 Campolo et al. 2017.
404 Marcus en Davis 2019: 34.
405 Hicks, 12 oktober 2018.

witte jongeren' foto's van vrolijke jongeren en 'drie Aziatische jongeren' foto's van schaars geklede meisjes.[406] Een ander voorbeeld komt van Schiphol. Een algoritme bedoeld om te helpen bij de logistiek rondom vliegtuigen op de baan, herkende een wit vliegtuig van Delta Airlines niet. Omdat het vooral getraind was op vliegtuigen van KLM, had het algoritme geleerd dat vliegtuigen altijd blauw waren. Allemaal voorbeelden waarin een algoritme de vooroordelen overneemt uit de trainingsdata.

Naast de vooroordelen in de trainingsdata, zorgen ook de keuzes in het ontwerp en de doelen van AI ervoor dat een systeem niet neutraal is. De eigenschappen of zienswijzen van de ontwikkelaars zelf kunnen bijvoorbeeld invloed hebben op het ontwerp van AI. Van verschillende gezichtsherkenningssoftware en van bepaalde automatische handzeepmachines is bekend dat zij de huid van zwarte mensen niet goed kunnen herkennen. Dat maakt duidelijk dat bij de ontwikkeling en het testen van deze applicaties met deze groep mensen geen rekening is gehouden. Broussard merkt op dat de bij de introductie van de Apple Watch allerlei gezondheidsdata gekwantificeerd konden worden, maar dat de ontwikkelaars niet hadden gedacht aan data over menstruatiecycli, voor vrouwen heel voor de hand liggende data.[407]

Zelfs als de data geen vooroordelen bevatten, kan de werking van een algoritme mensen benadelen door de keuze voor het doel waartoe het algoritme geoptimaliseerd wordt. Een algoritme in een ziekenhuis zou bijvoorbeeld geoptimaliseerd kunnen worden om zoveel mogelijk behandelingen te doen, om zoveel mogelijk geld te besparen of om de werkschema's van medisch personeel zo goed mogelijk in te richten. Op basis van dezelfde data ontstaan heel andere uitkomsten naar gelang de keuze van de doelen. Zoals Cathy O'Neil opmerkt, worden veel algoritmes niet ingezet om een veld te verbeteren, maar om besparingen door te voeren.[408] Achter de gepresenteerde objectiviteit kan dus een specifieke agenda schuilgaan.

Problemen met de keuze van doelen hoeven niet altijd te ontstaan door bewuste handelingen. Voor veel menselijke activiteiten geldt bijvoorbeeld dat zij tegelijkertijd verschillende doelen en belangen adresseren, waarvan sommige niet altijd expliciet en duidelijk zijn. Optimalisatie van het ene doel kan andere zaken in het gedrang brengen. Denk aan huisartsbezoeken, die dienen om de juiste diagnoses bij mensen te stellen. Een onlineplatform dat daarbij assisteert, kan de huisarts ontlasten zodat deze zich vooral met complexere ziektebeelden

406 Benjamin 2019: 93.
407 Broussard 2019: 157.
408 O'Neil 2016.

bezig kan houden. Maar voor sommige mensen geldt dat een bezoek aan de huisarts vooral dient om een geruststellende stem te horen of menselijk contact te hebben. Die implicietere doelen negeert een dergelijk platform.

Of neem de ogenschijnlijk eenduidige activiteit van navigeren. Algoritmes kunnen de snelste of kortste route van A naar B presenteren. Er zijn echter ook andere mogelijke doelen, zoals een bezoek aan een tankstation, het zoeken naar een mooier uitzicht, het vinden van een goede plek om onderweg te stoppen en te eten of het vermijden van kronkelige wegen. Navigatiediensten kunnen met veel van deze doelen rekening houden, maar het is duidelijk dat bij het kiezen van een route een veelheid van doelen in het spel kan zijn.[409]

De vooroordelen in trainingsdata en de keuzes die gemaakt worden tijdens het ontwerpproces zorgen er dus voor dat AI niet per definitie – of misschien zelfs per definitie niet – neutraal is. AI maakt het dan ook niet zonder meer mogelijk om processen te 'depolitiseren'. We zagen al dat met doelen verschillende waarden en agenda's gediend kunnen worden. Maar zelfs wanneer overeenstemming bestaat over het doel van een algoritme, is daarmee niet gezegd dat dat algoritme daar op neutrale wijze voor kan optimaliseren. Een algoritme kan middelen op heel verschillende manieren eerlijk verdelen.

Zo bestaat er een veelheid aan definities van eerlijkheid. Neem geslacht als variabele. Het is duidelijk dat er sprake is van discriminatie als daar bij een sollicitatie rekening mee wordt gehouden. Het is echter denkbaar dat wanneer een zwangerschap leidt tot een gat in het cv, juist wel rekening gehouden wordt met geslacht om mannen geen oneerlijk voordeel te geven. In andere gevallen vraagt het steunen van benadeelde groepen er juist om rekening te houden met bepaalde variabelen uit overwegingen van eerlijkheid. Een onderzoek heeft laten zien dat het wiskundig onmogelijk is om tegelijkertijd aan verschillende definities van eerlijkheid te voldoen.[410] Wiskundige algoritmiek kan politieke discussies over eerlijkheid dus niet vervangen.

Ook zijn vraagtekens te plaatsen bij de neutraliteit van AI vanwege het gebruik van allerlei *proxies*. Vaak is datgene wat we willen weten, moeilijk te berekenen of onduidelijk. Daarvoor worden dan andere variabelen gebruikt die indicaties moeten zijn voor datgene wat we willen meten. De architect Laura Kurgan formuleerde het nog sterker: "Wij meten de dingen die gemakkelijk [en] goedkoop te meten zijn."[411] De onlinewereld zit vol met *proxies* voor eigenschappen

409 Agrawal et al. 2018: 89.
410 Het AI-Now Institute verwijst hiervoor in het rapport uit 2018 onder andere naar Kleinberg 2018.
411 Greenfield 2017: 53.

van mensen: aantallen Facebookvrienden als maatstaf voor relaties tussen mensen, 'likes' als maatstaf voor populariteit, betaalgeschiedenis als maatstaf voor kredietwaardigheid. Zo zou een app van een Stanfordpromovendus meten of iemand een 'goede selfie' neemt. Dat zou gemeten worden op basis van objectieve standaarden, was de suggestie. Het algoritme was echter getraind op het aantal likes voor foto's op sociale media en mat dus eigenlijk populariteit. Als gevolg hiervan werden de selfies van jonge witte vrouwen standaard hoger beoordeeld en kreeg een oude zwarte man een lagere beoordeling, ongeacht op welke manier hij de selfie maakte.[412]

Ook op een breder maatschappelijk niveau speelt deze beperking. Steden worden bijvoorbeeld beoordeeld op basis van vage *'quality of life'*-indices, op de aanwezigheid van 'supercreatieve beroepen' en aantallen patenten dienen als indicator voor 'innovatiekracht'.[413] Het is daarom belangrijk te realiseren dat we vaak niet direct te maken hebben met het fenomeen waarin we geïnteresseerd zijn, maar dat het gebruik van *proxies* een vertekend en niet-objectief beeld kan creëren.

Proxies die nooit bewust geselecteerd zijn door softwareontwikkelaars, kunnen bovendien leiden tot vooroordelen in de algoritmes. Dit is een gangbaar probleem bij AI-systemen. Expliciet kunnen de makers ervan bepaalde variabelen als geslacht of etniciteit uit de dataset verwijderen. Desondanks kan een zelflerend algoritme zonder die data toch *proxies* voor die variabelen ontwikkelen en dan evengoed bepaalde groepen benadelen. Onderzoeken hebben bijvoorbeeld laten zien dat algoritmes het geslacht van sollicitanten kunnen bepalen op basis van woordgebruik. En postcodes kunnen als *proxy* dienen voor etniciteit. Binnen het veld wordt er veel onderzoek gedaan naar mogelijkheden om dit probleem met technische middelen te adresseren.[414]

Een vergelijkbaar bezwaar ten aanzien van de neutraliteit van AI heeft ermee te maken dat veel woorden die iets aanduiden dat voor ons van groot belang is, überhaupt geen objectieve betekenis hebben. Die zijn afhankelijk van onze keuzes en bestaan in feite per definitie uit subjectieve *'proxies'*. Een voor de hand liggend voorbeeld is 'schoonheid'. Het bedrijf Beauty AI ontwikkelde een app waarmee mensen hun foto konden versturen en dan beoordeeld werden op zogenoemd objectieve standaarden als symmetrie, rimpels en leeftijd. Kijkend naar de uitkomsten van deze schoonheidswedstrijd kwamen de makers erachter dat het algoritme mensen met een donkere huidskleur minder mooi vond.

412 Broussard 2019: 149.
413 Greenfield 2017: 56-57.
414 Van der Sloot et al. 2021a en 2021b.

Omdat schoonheid noodzakelijk subjectieve componenten bevat, zitten in de zogenoemde objectieve standaarden de subjectieve voorkeuren van de makers, of van de bevolkingsgroep of de sociale klasse waartoe de makers behoren.[415]

Maar denk ook aan een woord als 'gezondheid', waar veel AI op gericht is. Alhoewel er objectieve elementen in zitten, zijn er ook elementen waar mensen over van mening verschillen. Dat geldt ook voor woorden die weinig subjectiefs lijken te bevatten als 'armoede' en 'achterstandswijk', die echter wel de uitkomst zijn van politieke discussies en frames. Bovendien kan een algoritme een juiste voorspelling geven voor iets waarnaar gezocht wordt, maar daarbij in feite verwijzen naar een ander patroon. Zolang dat onderliggende patroon onzichtbaar blijft, kan de voorspelling onterecht als neutraal worden gepresenteerd. Een algoritme kan bijvoorbeeld terecht aangeven dat bepaalde mensen in contact met de politie zullen komen. Het is daarbij heel goed mogelijk dat het eigenlijk meet dat deze mensen uitgesloten worden door instituties en daardoor met de politie in aanraking komen. De achterliggende onrechtvaardigheid verdwijnt zo naar de achtergrond.

Zo werd de genoemde COMPAS-score bijvoorbeeld gebaseerd op een 137 punten tellende vragenlijst voor gearresteerden. Daarbij werd gekeken naar zaken als lage opleiding, financiële schulden, criminele vrienden en slechte gezinssituaties. In principe kon het algoritme aantonen dat deze zaken voorspellers zijn voor criminaliteit. Maar in plaats van criminele aanleg te meten, wat de suggestie is, zijn deze factoren indicatoren van armoede. Ze kwalificeren minder welvarende mensen daarmee als potentiële criminelen.[416] Dergelijke variabelen meten dus niet slechts criminaliteit, maar dragen ook bij aan de presentatie en productie ervan en zijn dus niet neutraal. In tegenstelling tot natuurwetenschappelijke methoden beïnvloedt veel AI-onderzoek zelf de uitkomsten: een kredietscore meet niet alleen de kans op een faillissement, maar vergroot die zelf ook.[417]

Een laatste fundamenteel probleem met de vermeende objectiviteit is dat hoe goed data ook mogen zijn, zij altijd slechts een bepaald aspect van de werkelijkheid weerspiegelen. Greenfield merkt in deze context dan ook op dat het woord 'data' misleidend is. 'Data' is namelijk Latijn voor 'gegeven'. Beter zouden we volgens Greenfield kunnen spreken van 'capta': datgene wat genomen wordt.[418] Data geven grip op iets. Daarmee brengen ze altijd een machtsdimensie met zich

415 Benjamin 2018: 49-51.
416 Broussard 2019: 156.
417 Pasquale 2015: 41.
418 Greenfield 2017: 210.

mee. Ze structureren wat gemeten wordt en wat niet, categoriseren en volgen indelingen. Denk in het laatste geval aan een binaire indeling in gender die voor bepaalde groepen mensen niet adequaat is. Ook zoiets als de *Quantified Self Movement,* waarbij middels *wearables* allerlei data wordt verzameld, doet het menselijk lichaam op een bepaalde manier verschijnen en suggereert manieren om er met een fitnessregime controle over te krijgen. Het is belangrijk zich bewust te zijn van die machtsdimensie, zeker wanneer geclaimd wordt dat een algoritme volstrekt neutraal is.[419]

De conclusie van bovenstaande bezwaren tegen het idee dat AI neutraal is, is niet dat het gebruik ervan afgeraden moet worden of dat AI nooit neutraler kan zijn dan mensen. Dat kan ze zeker wel. Wat de bezwaren aantonen, is hoe complex het gebruik van AI is en waar allemaal op gelet moet worden als we de technologie ergens voor willen inzetten. Ze wijzen ons op de vragen, technische uitdagingen en discussies die met een verantwoord gebruik van AI gepaard moeten gaan. Als wij daar niet op letten en blind vertrouwen op het neutrale oordeel van algoritmes, kunnen daardoor allerlei misstanden ontstaan en vooroordelen gecodeerd worden, terwijl de suggestie is dat er een onbevooroordeelde behandeling plaatsvindt. Aan het eind van deze paragraaf presenteren we een lijst met vragen voor AI-systemen die volgen uit de drie besproken mythen over de werking van AI.

Kernpunten – Beeld 'AI is neutraal'

- Omdat AI gevoelens en andere menselijke eigenschappen mist, wordt gesuggereerd dat de technologie neutraal is, terwijl de werking ervan wel degelijk vooroordelen en misstanden kan bevatten
- De volgende factoren plaatsen vraagtekens bij de neutrale werking van algoritmes: De kwaliteit van de trainingsdata, eigenschappen van de ontwikkelaars van de technologie, de doelen waarvoor algoritmes ingezet worden, tegenstrijdige definities van eerlijkheid, het (onbedoeld) gebruik van *proxies*, de subjectieve betekenis van woorden en het filterende karakter van data en de macht die dat met zich meebrengt
- Dit zijn geen argumenten tegen het gebruik van AI, maar voor het stellen van vragen teneinde de technologie op verantwoorde wijze te gebruiken.

419 Zoals de zeventiende-eeuwse Franse kardinaal Richelieu zei: "geef mij zes regels geschreven door de hand van een van de meest eerlijke mensen en ik zal daarin iets vinden om hem mee op te hangen" (Greenfield 2017: 62).

'Kunstmatige intelligentie kent een bovenmenselijke rationaliteit'

Nauw verbonden met het beeld dat AI neutraal is, is de gedachte dat het rationeel is, of in ieder geval aanzienlijk rationeler dan mensen. Neutraliteit suggereert dat de uitkomsten eerlijker zijn, rationaliteit dat zij op superieure data en rekenkracht zijn gebaseerd en dat op grond daarvan patronen en verbanden geïdentificeerd kunnen worden die voor mensen te complex zijn.

Die rationaliteit is een grote belofte van veel AI-toepassingen. Neem het domein van de zorg. Bij het stellen van een diagnose vergelijkt een arts de data die voorliggen, met eerder opgebouwde kennis en ervaring. De mens is daarin noodzakelijkerwijs beperkt. Bovendien groeit de hoeveelheid kennis in een hoog tempo. Volgens schattingen zou een medisch specialist het grootste deel van de tijd bezig moeten zijn met het lezen van onderzoek om op de hoogte te blijven van de nieuwste kennis in het eigen veld. Dat is een onmogelijke opgave. Zeldzame genetische aandoeningen, die bijvoorbeeld voorkomen bij een populatie die vooral uit het buitenland afkomstig is, zijn dan ook lastig te diagnosticeren. AI daarentegen kan immense databases bestuderen en steeds geüpdatet worden met de nieuwste medische kennis. Dat was ook de belofte toen IBM's Watson in de gezondheidszorg werd ingezet.

We zullen in andere hoofdstukken zien waarom dit zo simpel nog niet is. Maar de logica erachter lijkt duidelijk: AI-systemen kunnen veel meer data verwerken dan mensen en hebben immense rekenkracht tot hun beschikking. De beslissingen die zij nemen, kunnen daarom als rationeler en nauwkeuriger beschouwd worden. Ook prominente onderzoekers van AI kunnen een naïef geloof hebben in de manier waarop deze technologie meer rationaliteit zal brengen in bijvoorbeeld politiewerk of de werking van financiële markten.[420]

Daar kunnen wij echter vier soorten kanttekeningen bij plaatsen. Ten eerste meten veel AI-systemen correlatie en dat is wat anders dan causaliteit. Al eeuwen geleden heeft de wetenschapsfilosofie het verschil tussen beide fenomenen laten zien, maar in de praktijk worden die toch regelmatig met elkaar verward. Dat twee fenomenen regelmatig samen voorkomen, betekent niet dat de een de oorzaak is van de andere. Er kan bijvoorbeeld sprake zijn van een veel complexere causaliteit of van toeval. Een voorbeeld van het eerste was een schaakprogramma dat een patroon destilleerde waarbij het weggeven van een koningin vaak voorafging aan het winnen van een partij. Het programma identificeerde dat dus als een goede zet. Het offeren van de koningin is echter

420 Agrawal et al. 2018: 75; Domingos 2015: 20, 43.

zeer kostbaar en werd gedaan omdat in die context een grotere prijs, schaakmat, mogelijk was.[421]

In de tweede plaats is de gepresenteerde rationaliteit van een AI-systeem vaak onderdeel van de promotie van bepaalde diensten en producten waarmee menselijke rationaliteit onterecht in een negatief daglicht wordt geplaatst. Broussard geeft een voorbeeld uit het domein van mobiliteit. Ze merkt op dat rondom zelfrijdende auto's een bepaald feit zeer vaak terugkomt, namelijk dat wereldwijd 1,2 miljoen mensen sterven in auto-ongelukken, in 95 procent van de gevallen door menselijke fouten. Dat klinkt als een sterke motivatie om mobiliteit te automatiseren. Maar, zoals Broussard terecht opmerkt, het is ook logisch dat bijna alle ongelukken het resultaat zijn van menselijke fouten. Mensen zijn namelijk de enige bestuurders van auto's. Het zou vreemd zijn als de data anders waren.[422] Een ander voorbeeld van de wijze waarop het vermogen van de mens negatief wordt afgezet tegen dat van AI, is de terminologie rondom datatoepassingen zoals IBM's Watson. Daarbij wordt vaak gesproken over 'dark data', ofwel data die nog niet 'digitaal gevangen' (digitaal beschikbaar) zijn, een term die associaties opwekt met gebrek aan controle, weerbarstigheid en rebelsheid. Door bestaande praktijken als donker neer te zetten, worden digitale oplossingen gepresenteerd als brengers van transparantie en rationaliteit. Zij redden ons van de verspilling van allerlei data.[423] Dergelijke frames zijn niet beperkt tot AI en behoren tot grotere ideologische standpunten die Broussard duidt als 'techno-chauvinisme' en Zuboff als de retoriek van het 'surveillance-kapitalisme'. Later in dit hoofdstuk komen we hierop terug als we bredere technologiebeelden bespreken. Hier is het belangrijk op te merken dat de grotere rationaliteit van AI ook onderdeel kan zijn van onrealistische voorstellingen van de wereld.

Een derde kanttekening bij het idee dat AI rationeel is, heeft te maken met een dynamiek die we ook bij eerdere systeemtechnologieën zijn tegengekomen, namelijk de misleidende kracht van woorden. Zoals we in hoofdstuk 1 beschreven, neigen we ernaar AI te begrijpen in menselijke termen. Anders gezegd, we proberen haar te antropomorfiseren. Herinner de Moravec-paradox. Voor een mens is schaken iets dat veel intelligentie vereist. Als een machine dat kan, zijn we geneigd die prestatie te extrapoleren naar bredere intellectuele vermogens, terwijl dat niet terecht is. Dat een machine ons in schaken overtreft, is vergelijkbaar met een paard dat sneller beweegt dan wij. Daarmee kunnen beide ons nog

421 Kasparov 2018: 99-100.
422 Broussard 2019: 136-137. Dit veelgebruikte datapunt traceert zij dan ook tot een maker van autonome voertuigen voor het leger.
423 Zuboff 2019: 210-211.

niet op andere domeinen overtreffen. Het is dan ook belangrijk om te erkennen op welke manier machines anders in elkaar zitten dan het intelligent gedrag van mensen.

Een vierde kanttekening bij de rationaliteit van AI-systemen verdient bijzondere aandacht, omdat deze een groeiend fenomeen betreft met potentieel kwalijke effecten. Dat is namelijk de aandacht voor AI vanuit pseudowetenschappelijke theorieën en toepassingen. Een van de meest prominente voorbeelden daarvan is het veld van emotiedetectie. Dat is een onderdeel van gezichtsherkenning, waarbij de claim is dat de achterliggende emoties uit gezichten gedestilleerd kunnen worden. Het bedrijf Kairos claimt bijvoorbeeld woede, vrees en verdriet uit videobeelden te kunnen afleiden. In 2019 gaf Amazon aan dat het systeem Rekognition acht verschillende emoties van gezichten af kon lezen. Het herkennen van emoties wordt bijvoorbeeld gebruikt bij sollicitaties, waar een bedrijf als HireVue diensten voor aanbiedt. In China wordt emotiedetectie gebruikt om te kijken of leerlingen in de klas wel aan het opletten zijn. Ook het Amerikaanse bedrijf BrainCo werkt hieraan. Uit stemgeluid destilleren programma's als Cogito en Empath de emoties van mensen die callcenters bellen. Veiligheidsdiensten in de VS en het Verenigd Koninkrijk gebruiken deze toepassing om te onderzoeken of mensen liegen of iets verbergen. Dit is een groeiende toepassing van AI: de waarde ervan zal volgens een schatting stijgen van 12 miljard dollar in 2018 naar 90 miljard dollar in 2024.[424]

De industrie groeit dus, maar er is geen wetenschappelijke basis voor het veld van emotiedetectie. De oorsprong ervan ligt in het werk van de psycholoog Paul Ekman in de jaren zestig. Hij ontwikkelde een methodiek om 27 'actie-units' in gezichten te onderscheiden en concludeerde dat er zes basisemoties zijn. Het veld is op zijn werk gebaseerd, maar er is nog geen bewijs dat dit correct is.[425] Sterker nog, er is aanleiding om te geloven dat het voelen en uiten van emoties varieert tussen culturen, individuen en zelfs bij een enkel persoon door de tijd heen. Het zorgelijke is dus dat terwijl de kennis hierover dubieus is, er wel naar gehandeld wordt – waardoor kinderen gestraft worden als ze niet opletten[426], en mensen afgewezen worden voor banen of verdacht worden van leugens.

Emotieherkenning is een voorbeeld van een breder fenomeen van op algoritmes gebaseerde pseudowetenschap. Naast emotiedetectie wordt bij sollicitaties ook gebruikgemaakt van onlinepersoonlijkheidstesten om te onderzoeken of

424 Crawford et al. 2019.
425 Zuboff 2019: 285.
426 In 2018 werd als reactie op emotiedetectie in de klas in China de hashtag #ThankGodIGraduatedAlready trending (Pasquale 2020: 60).

iemand past bij de een specifieke baan. Ook op dit gebied geldt dat er geen bewijs is voor een duidelijke indeling in persoonlijkheidstypen die voorspellende kracht over iemands vaardigheden heeft.[427]

Een ander voorbeeld zijn allerlei fitnesstrackers en wearables. Er is veel twijfel over de accuraatheid van de metingen van beweging, de verbranding van calorieën of de duur van iemands slaap. Tegelijkertijd zien veel mensen deze toepassingen wel als een wetenschappelijke manier om hun gezondheid in kaart te brengen.

Een eveneens dubieuze toepassing is het gebruik van gezichtsherkenningssoftware om de seksuele geaardheid van mensen te bepalen. Onderzoekers beweerden dat met grote accuraatheid te kunnen doen.[428] Een belangrijk risico hierbij is dat homoseksualiteit in veel landen iets strafbaars is. Zelfs als de software accuraat zou zijn – wat allerminst duidelijk is –, dan is deze in de handen van autoritaire regimes voor veel mensen een groot gevaar.

Wat zijn nu mogelijke verklaringen voor het feit dat, ondanks de grote hoeveelheden data en rekenkracht, AI toch gebruikt kan worden voor praktijken die gestoeld zijn op dit soort pseudowetenschappelijke theorieën? Een reden is dat het hier thematiek betreft die wij amper begrijpen, die zeer complex is en daarmee moeilijk te testen of te weerspreken is. Hoe bewijs ik bijvoorbeeld dat ik geen ongeduldige persoonlijkheid heb? Of dat ik toch niet voldoende slaap heb gehad? Dat ik wel aan het opletten was in de klas? Zaken als seksuele geaardheid zijn zeer complex en kunnen niet volledig in een binair onderscheid tussen heteroseksueel en homoseksueel gevangen worden, zoals de grote diversiteit binnen de LGBTQIA-gemeenschap aantoont. Dergelijke analyses zijn dus onwetenschappelijke simplificaties.[429]

Een andere oorzaak van het gebruik van pseudowetenschappelijke theorieën is dat er simpelweg geen terugkoppeling plaatsvindt om te testen of de voorspelling correct was of niet. Van iemand die afgewezen wordt voor een baan op basis van een persoonlijkheidstest kunnen we nooit weten of hij of zij eventueel toch geschikt zou zijn geweest. Iemand wiens asielaanvraag wordt afgewezen omdat hij of zij gelogen zou hebben, heeft vaak geen mogelijkheid om de eigen onschuld te bewijzen. Dergelijke technologische toepassingen onderzoeken

427 O'Neil 2016.
428 Claus, 12 september 2017.
429 Een heel vergelijkbaar fenomeen is het gebruik van DNA-tests door de Britse immigratiedienst om nationaliteit te bepalen en daarmee bijvoorbeeld Kenianen van Somaliërs te onderscheiden. Genen houden zich echter niet aan nationale grenzen en de gedachte dat nationaliteit biologisch gemeten kan worden, is incorrect (Benjamin 2019).

dus niet alleen een bepaald gebied, maar produceren zelf ook een werkelijkheid, vaak zonder getoetst te worden.

Kernpunten – Beeld 'AI is bovenmenselijk rationeel'

- Meer data, grotere rekenkracht en het vermogen om complexe verbanden te onderscheiden suggereren dat AI bovenmenselijk rationeel is.
- Correlatie wordt echter vaak verward met causaliteit.
- Sommige presentaties van de vaardigheden van AI dienen commerciële doelen.
- Antropomorfisering van AI wekt de indruk van grotere intellectuele vermogens dan daadwerkelijk het geval is.
- Pseudowetenschappelijke theorieën en toepassingen worden gebruikt in de vorm van emotiedetectie, onlinepersoonlijkheidstesten, fitnesstrackers, analyses van seksuele geaardheid en bepaalde wetenschappelijke benaderingen van armoede. Dit brengt grote maatschappelijke risico's met zich mee.

'Kunstmatige intelligentie is een black box'

Een veelgehoorde karakterisering van AI is dat het een zogenoemde black box is. Die term is vooral gepopulariseerd door de vroege cybernetici en doelt op systemen die wij niet goed kunnen doorzien en begrijpen. Een input wordt door een black box op mysterieuze wijze vertaald in een output zonder dat we de interne werking ervan bevatten.[430] AI zou daardoor niet transparant, onuitlegbaar en nauwelijks te reguleren zijn. Dat is met name problematisch voor domeinen waar transparantie juist belangrijk is, zoals de motivatie om tot de veroordeelding van een verdachte te komen. Maar ook in andere gevallen waar legitimiteit, rechtszekerheid en rechtsgelijkheid cruciaal zijn, speelt het probleem van de black box. Illustratief in dit verband is de opstelling van zowel de Nederlandse bestuursrechter als de Raad van State om meer transparantie bij de inzet van algoritmes te vereisen.[431] In gevallen die minder van 'levensbelang' zijn, wordt vaak betoogd dat controle en transparantie niet nodig zouden zijn, maar op de lange termijn kan dat toch problematisch zijn.

Vanwege de associatie van AI met een black box spreekt Frank Pasquale zelfs van de opkomst van een 'Black Box Society'. Typerend voor deze samenleving

430 Rid 2016: 66-67.
431 Wolswinkel 2019: 776-785.

is dat er in bijvoorbeeld reputatiesystemen, zoekmachines en de financiële sector beslissingen worden genomen door zwarte dozen. Pasquales gebruik van de term kent twee dimensies. Niet alleen gaat het om onbegrijpelijke systemen, maar ook, net als bij de zwarte doos van een vliegtuig, om een opnameapparaat.[432]

Is het een mythe dat AI een black box is? Niet zonder meer. Maar het is wel belangrijk om preciezer te zijn met wat we ermee bedoelen. De term 'black box' wordt vaak op heel verschillende manieren gebruikt. Sommige daarvan gaan met grotere obstakels om ze te adresseren gepaard dan andere. We moeten die verschillende betekenissen onderscheiden om adequate antwoorden op het vraagstuk te kunnen formuleren.

Het idee van een black box wordt allereerst gebruikt om aan te geven dat iets zeer complex is en dat bepaalde mensen het daarom niet begrijpen. Dat betekent niet dat er niet andere mensen zijn die het wel begrijpen. In deze zin is een groot deel van de moderne samenleving een zwarte doos voor de meeste mensen. Wanneer ze in een lift stappen, vertrouwen ze daarop zonder te weten hoe het werkt. Dit geldt niet alleen voor technische zaken, maar ook voor veel juridische, politieke en bestuurlijke zaken. Dat betekent echter niet dat de zaken niet te begrijpen zijn. Er zijn groepen mensen die zich in dit soort zaken bekwamen en er verantwoordelijkheid voor dragen.

De analogie bij AI voor deze vorm van een black box is de beslisboom van een expertsysteem. Die is moeilijk te begrijpen voor iemand die daar niet in geschoold is, maar kan door iemand die dat wel is goed uitgelegd worden. Deze vorm levert niet veel problemen op en vraagt alleen dat wij ervoor zorgdragen dat er mensen zijn die het systeem begrijpen en kunnen uitleggen, net zoals er altijd voldoende monteurs moeten zijn om een defecte lift te repareren.

Een tweede manier waarop van een black box sprake kan zijn, is wanneer we geen toegang hebben tot de data en analyses op basis waarvan uitkomsten gegenereerd zijn. Dit kan verschillende oorzaken hebben. Een eerste mogelijkheid is dat die data simpelweg niet zijn bijgehouden of opgeslagen. Het kan ook zijn dat iemand geen recht heeft om die informatie in te zien, bijvoorbeeld wanneer een organisatie gebruik maakt van de diensten van een bedrijf dat de data en de werking van het algoritme als een bedrijfsgeheim beschouwt. Of als een overheid een algoritme niet openbaar wil maken, omdat anders het doel ervan (zoals fraudebestrijding) wordt ondermijnd. In deze variant is sprake van een black box omdat een bepaalde partij de feitelijke mogelijkheid niet heeft

---

432 Pasquale 2015.

om het systeem te begrijpen. De black box is hier veelal van commerciële en juridische aard. Bijvoorbeeld omdat contractuele afspraken tot geheimhouding zijn gemaakt of de intellectuele eigendomsrechten toegang in de weg staan. Ook deze variant van het idee van een black box is niet onoverkomelijk.

Pasquale betoogt bijvoorbeeld dat we kritisch moeten kijken naar allerlei wetgeving die het in de VS gemakkelijker heeft gemaakt om zaken als bedrijfsgeheimen te classificeren, zeker omdat het gebruik van bedrijfsgeheim nu grote delen van de samenleving beïnvloedt.[433] Ook het Amerikaanse AI Now Institute betoogt dat we niet moeten accepteren dat de werking van belangrijke systemen in de samenleving bedrijfsgeheimen zijn.[434] Het gaat hierbij niet alleen om regels van geheimhouding, maar ook om contractuele bepalingen die geweigerd kunnen worden. Een lastiger variant van dit gebruik van de term 'black box' is wanneer de data waar een uitkomst op gebaseerd is, verspreid zijn over hele verschillende bronnen.[435] Bijvoorbeeld algoritmes die gebaseerd zijn op andere algoritmes waarvan de herkomst niet te traceren is. Dit probleem speelt bij ketenbesluiten. Het onderzoek van Marlies van Eck laat zien hoe binnen de overheid besluiten worden genomen door allerlei systemen aan elkaar te koppelen.[436] Alhoewel de uitkomst in principe niet onbegrijpelijk is, is het in de praktijk nagenoeg onmogelijk om na te gaan hoe de besluitvorming tot stand is gekomen.

Een derde gebruik van de term 'black box' is technischer van aard en hangt nauw samen met de recente opkomst van *deep learning* in AI. Daarbij gaat het om systemen die qua complexiteit zo geavanceerd zijn dat het voor mensen te lastig zou zijn om de uitkomsten te begrijpen. Neem het wel of niet plaatsen van een bepaald artikel op iemands tijdlijn op Facebook. Dat is een immens complex proces dat in realtime plaatsvindt met miljoenen gebruikers tegelijkertijd waarvan de data met elkaar interacteren. Dit vraagstuk speelt bijvoorbeeld bij de vrees voor beïnvloeding van verkiezingen. Het gevaar is dat het gegeven de complexiteit misschien niet meer mogelijk is om te achterhalen waarom een bericht wel of niet op de tijdlijn van iemand verscheen.[437] Belangrijk in deze categorie is dat hiermee niet gezegd is dat dat principieel onbegrijpelijk is. Het achterhalen van hoe een systeem tot een beslissing komt is voor onderzoekers alleen zeer complex.

433 Pasquale 2015.
434 Crawford et al. 2019.
435 WRR 2011.
436 Van Eck 2018.
437 Greenfield 2017: 252-253.

Soms zijn de processen van een systeem niet alleen zeer complex, maar *te* complex voor een mens om na te gaan. Dat komt doordat een AI-systeem soms een andere logica volgt dan die van het menselijk brein. Denk bijvoorbeeld aan beeldherkenning op het niveau van pixels. We kunnen best begrijpen hoe een gezicht herkend wordt op basis van een deel van een foto, maar deep learning onderscheidt patronen op verschillende diepere lagen en gebruikt zo input op het niveau van individuele pixels. Een redenering op dat niveau is voor de mens onnavolgbaar. Bij Facebook zouden twee computerprogramma's een 'taal' ontwikkeld hebben waarmee ze met elkaar konden communiceren op een manier die mensen niet begrijpen. Het is in dit soort gevallen echter wel altijd de vraag of hier sprake is van principiële onbegrijpelijkheid of dat we een proces op den duur wel degelijk kunnen vatten.

Hier gaan we niet in op wat specifieke remedies zijn voor de verschillende vormen die black box kan hebben. Wat we willen laten zien, is dat de term voor heel verschillende fenomenen wordt gebruikt en dat dat tot verwarring leidt. Sommige van die fenomenen brengen grotere obstakels mee dan andere. Onvermogen om te begrijpen wordt niet alleen veroorzaakt door de aard van algoritmes, maar ook door eigendomsrechten of complexe sociale systemen. Het antwoord op hoe hiermee om te gaan, verschilt dan ook bij deze vier vormen. Wanneer de term 'black box' wordt gebruikt om aan te geven dat het niet mogelijk is iets te begrijpen en we het daarom niet zouden moeten gebruiken of het niet goed kunnen reguleren, is het tijd om te pauzeren en de mythe door te prikken.

Kernpunten – Beeld 'AI is een black box'

- Het beeld van AI als black box kan leiden tot het idee dat controle of transparantie onmogelijk is.
- De term 'black box' wordt echter op heel verschillende manieren gebruikt: als aanduiding van complexiteit (die door bepaalde groepen mensen wel goed te begrijpen is), als afwezigheid van (juridische) toegang tot de werking van een systeem, als aanduiding van een grote hoeveelheid berekeningen en als fundamentele onbegrijpelijkheid.
- Omdat met het gebruik van de term 'black box' naar verschillende zaken verwezen kan worden, is het van belang om helder te hebben wat wij eronder verstaan.

Veel misverstanden over de werking van AI zijn tegen te gaan door kritische vragen te stellen tijdens de verschillende stappen bij het toepassen van AI. Het kader op de volgende pagina doet daarvoor een aantal suggesties.

Vragen om te stellen bij AI in de praktijk

Opzet & planning

- Welke uitsnede van de werkelijkheid wordt er gemaakt?
- Is de bestaande praktijk goed geanalyseerd?
- Welk doel wordt door het systeem geoptimaliseerd?
- Dient het toepassingsdomein meerdere doelen?
- Hoe beïnvloedt het wereldbeeld van de makers het systeem?
- Zijn er mensen die de werking van het algoritme kunnen uitleggen?
- Is er toegang tot de databases en modellen waarmee het algoritme is getraind?
- Zijn er juridische obstakels voor inzage in de werking van een algoritme?

Data-verzameling & training

- Wat is de kwaliteit van de trainingsdata?
- Wordt het fenomeen direct gemeten of worden proxies gebruikt?
- Wordt überhaupt een objectief fenomeen gemeten?
- Schuilen er pseudowetenschappelijke theorieën achter het model?

Werking van het systeem

- Welke definitie van eerlijkheid wordt gehanteerd?
- Is er sprake van patronen die voor het menselijk brein niet te begrijpen zijn?
- Worden meerdere databronnen gebruikt die inzicht bemoeilijken?

Output & effect

- Wordt correlatie onderscheiden van causaliteit?
- Suggereren bepaalde woorden een onterechte analogie met menselijke intelligentie?
- Beïnvloedt het algoritme ook wat het meet?

Mythen over de consequenties van AI

'Kunstmatige intelligentie zal binnenkort de mens evenaren'

In het voorgaande hebben we drie mythen besproken over hoe AI werkt: dat de technologie neutraal, rationeel en een black box zou zijn. Vervolgens kijken we naar de mythen over de verwachte implicaties van AI in de nabije toekomst. Het verhaal van de robot Sophia (zie tekstbox 4.1) is illustratief voor zo'n mythe, namelijk dat we binnenkort kunstmatige intelligentie kunnen verwachten die de mens evenaart en ons vervolgens ver zal overstijgen.[438] Zoals we in deel 1 van dit rapport hebben gezien, wordt al sinds het begin van het veld gespeculeerd over dit soort *artificial general intelligence* (AGI).

Tekstbox 4.1 – Robotburgers
In 2016 werd de robot Sophia, ontwikkeld door het bedrijf Hanson Robotics, getoond op het beroemde technologiefestival South by Southwest. Een jaar later stond zij op de *Future Investment Summit* in Riyad, waar zij het staatsburgerschap van Saoedi-Arabië kreeg. "Het is historisch om de eerste robot te zijn die het burgerschap verleend krijgt", antwoordde zij zelf. Onmiddellijk ontstonden er allerlei discussies over wat dit betekende. Had Sophia hiermee het recht om te trouwen of te stemmen? En zou haar uitzetten dan nu een daad van onrecht betekenen?

De wiskundige Good sprak over een toekomstige *"intelligence explosion"* en de schrijver Vernor Vinge muntte de term *"singularity"* voor het moment waarop volgens hem slimme machines zich tot ons zullen verhouden zoals wij tot ons tot dieren verhouden.[439] Dergelijke verwachtingen zijn de laatste jaren weer nieuw leven ingeblazen. Futurist Ray Kurzweil, werkzaam voor Google, verwacht dat AGI in 2029 zal bestaan en dat de singulariteit rond 2045 zal plaatsvinden. DeepMind heeft, evenals andere bedrijven als Vicarious, Kindred en Numenta, het creëren van AGI expliciet als missie geformuleerd.[440]

438 Een vraagstuk dat hierop voortborduurt, betreft het effect van AI op de werkgelegenheid en de angst voor massawerkloosheid. Alhoewel bij eerdere systeemtechnologieën die angst een mythe bleek te zijn, is het in de context van AI ook een belangrijk vraagstuk. Dit vraagstuk komt aan de orde in het volgende hoofdstuk, waar we kijken naar de inbedding van AI in de macro-economische context.
439 Vinge 1993.
440 Agrawal et al. 2018: 223.

De verwachting dat AI binnenkort menselijke vermogens kan evenaren, wordt ingegeven door recente ontwikkelingen in AI en de suggestie dat er allerlei doorbraken richting AGI plaatsvinden. Op het gebied van de zelfrijdende auto slaagde Otto, een divisie van Uber, erin om een auto zelf van de oostkust naar de westkust van de VS te laten rijden. In een interview in 2016 merkte ook president Obama over zelfrijdende auto's op: "de technologie is er in principe al".[441]

Het idee dat er fundamentele doorbraken plaatsvinden, wordt zoals we aan het begin van dit hoofdstuk zagen ook gevoed door allerlei competities die met veel publiciteit gepaard gaan. De retoriek die daarbij gebruikt wordt, stuurt vaak in de richting van onterechte extrapolaties. Elk evenement wordt gepresenteerd als weer een stapje dichter bij het evenaren van alle menselijke intellectuele vaardigheden. Melanie Mitchell noemt dit een van de valkuilen waarin we neigen te trappen bij het denken over AI: dat *narrow intelligence* op een continuüm ligt met *general intelligence*.[442] Watson won dan wel met *Jeopardy!* maar is nog steeds geen goede arts: in 2017 beëindigde MD Anderson Cancer Center de samenwerking met IBM's Watson omdat sommige aanbevelingen ervan "onveilig en incorrect" waren.[443]

De geschiedenis van systeemtechnologieën leert ons terughoudend te zijn met verwachtingen rondom dit soort competities en demonstraties. Ze genereren aandacht, spreken tot de verbeelding, maar zijn er ook om de technologie te promoten en verhullen vaak sterk de gecontroleerde omstandigheden waarin ze plaatsvinden. De indrukwekkende rit van Otto vond bijvoorbeeld plaats in een zeer gecontroleerde omgeving. In meer alledaagse en derhalve aanzienlijk minder gecontroleerde situaties hebben al verschillende dodelijke ongelukken met zelfrijdende auto's plaatsgevonden sinds de eerste keer in 2016. Steeds meer grote obstakels voor zelfrijdende auto's komen aan het licht. Zo kunnen mensen gemakkelijk objecten in hun hoofd roteren en verplaatsen, maar voor algoritmes is dat moeilijk. Het niet herkennen van een omgevallen oranje pion zou daarmee gevaarlijke situaties op kunnen leveren. In het volgende hoofdstuk gaan we uitgebreider in op de stand van zaken rondom de zelfrijdende auto.

Over de zelfrijdende auto wordt al zeker sinds 2012 gezegd dat die er over een paar jaar zal zijn, maar die tijdlijn wordt steeds verder vooruit geschoven Dat zet de verwachtingen rondom AI in perspectief. Marcus en Davis merken ook op dat we hoopten op Rosie, de robotbediende uit de tekenfilmserie *The Jetsons,*

441 Broussard 2019: 142-147.
442 Mitchell 2021.
443 Marcus en Davis 2019: 5.

en in plaats daarvan kregen we Roomba, de autonome stofzuiger.[444] Zelfs de technologieondernemer Peter Thiel zei: "We wilden een vliegende auto, maar we kregen 140 karakters", verwijzend naar de ruimte voor een twitterbericht.

Een zeer populaire vorm van competitie om innovatie aan te jagen zijn de zogenoemde *hackatons*, die vanuit Silicon Valley de wereld over zijn gegaan. Bekenden in het veld geven echter aan dat de bravoure waarmee daarbij aankondigingen wordt gedaan, met een korrel zout genomen moet worden. Dergelijke trajecten zijn te kort om echt stappen te zetten richting levensvatbare producten. De producten van hackatons worden dan ook weleens schertsend aangeduid als 'vaporware', grote beloften van een innovatie die er nooit zal komen.[445]

Veel recente doorbraken zijn relevant, maar moeten wel in de juiste context geplaatst worden. Voor het spel *Jeopardy!* geldt bijvoorbeeld dat bijna 95 procent van de antwoorden de titels zijn van Wikipediapagina's.[446] Het winnen ervan demonstreert dus de vaardigheid om daar doorheen te navigeren, maar niet de beheersing van de complexiteit van menselijke taal. Volgens de filosoof Daniel Dennett zijn zelfs de spelregels van *Jeopardy!* nog wat verscherpt om Watson mee te kunnen laten spelen.[447] Van de overwinning op een schaakgrootmeester merkten we reeds op dat dit gebeurde volgens een lineair pad van vooruitgang vanaf de jaren zestig.[448] Voor het spel Go is enorm veel rekenkracht nodig en de overwinning op Lee Sedol was indrukwekkend. Tegelijkertijd was een combinatie aan methoden en het inprogrammeren van de kennis van een groot aantal menselijke experts nodig om het algoritme dit te kunnen laten doen.[449] Alhoewel het veel complexer is, is het type vraagstuk van Go vergelijkbaar met het spel boter-kaas-en-eieren, namelijk een bordspel op een gridpatroon waarvan de optimale uitkomst als een functie uitgedrukt kan worden.[450] Het winnende programma AlphaGo kan voor weinig anders dan dit type vraagstukken ingezet worden.

Demystificatie van AI betekent dat we onrealistische verwachtingen moeten adresseren. Alhoewel belangrijke stappen worden gezet, zijn we niet dicht bij het evenaren van de mens, AGI en het overschaduwd worden door AI. Sommige mensen gaan echter erg ver in het bagatelliseren van de kans op bovenmenselijke intelligentie. Volgens Andrew Ng is die zorg "vergelijkbaar

444 Marcus en Davis 2019: 98.
445 Broussard 2019.
446 Broussard 2019: 82.
447 Dennett 2019: 49.
448 Zie hoofdstuk 2.
449 Echter wel in contrast met AlphaZero, dat zichzelf trainde.
450 Broussard 2019: 33-4.

met de zorg voor overbevolking op Mars".[451] Stuart Russell plaatst echter terecht vraagtekens bij dat argument. Waar we namelijk nog niet bezig zijn om Mars te kolonialiseren, wordt nu al wel degelijk flink geïnvesteerd in de ontwikkeling van AGI.[452] Het bereiken daarvan is uiteindelijk het doel van het AI-vakgebied. Volgens Russell is het dan nogal vreemd om, terwijl we een trein aan het ontwikkelen zijn die op een afgrond afgaat, te zeggen dat we ons daar geen zorgen over hoeven te maken omdat we voor de tijd dat de afgrond in zicht komt, toch wel door onze brandstof heen zullen zijn.

Het evenaren van menselijke intellectuele vermogens dienen wij als doel dus wel serieus te nemen. Tegelijkertijd moeten we aankondigingen van doorbraken wel in een context plaatsen en zien dat dat doel met de huidige stand van zaken nog ver buiten ons bereik ligt. Russell geeft een mooie indeling van verschillende variabelen waarmee we naar AI-toepassingen kunnen kijken. De omgeving ervan kan volledig overzichtelijk zijn – zoals een schaakbord – of niet – zoals het wegverkeer –, handelingen kunnen discreet zijn of continu, er kunnen andere actoren zijn of niet, de uitkomsten van handelingen kunnen voorspelbaar zijn of niet, de omgeving kan dynamisch veranderen of niet en de horizon waarop het bereiken van doelen wordt gemeten kan kort of lang zijn.[453] Deze variabelen leveren een grote set aan verschillende soorten vraagstukken op. Terwijl grote vooruitgang wordt geboekt op vraagstukken die bijvoorbeeld volledig overzichtelijk, discreet en voorspelbaar zijn, is er nog een lange weg te gaan om andere typen vraagstukken op te lossen.

Veel van de vraagtekens die wij geplaatst hebben bij de verwachting dat de mens op korte termijn door AI geëvenaard wordt, vallen onder de drie elementen van wat Marcus en Davis de "AI-kloof" tussen verwachting en realiteit noemen. Het eerste element is onze goedgelovigheid. We schrijven aan machines zaken toe die voor onszelf gelden. Terwijl een bepaalde handeling uitvoeren van mensen intelligentie vraagt, hoeft dat voor machines niet het geval te zijn. Het tweede aspect van de kloof gaat over denkbeeldige vooruitgang. Vooruitgang op simpele problemen (zoals *Jeopardy!*) moet niet verward worden met vooruitgang op moeilijke problemen (zoals menselijke taal begrijpen). Ten slotte noemen zij een kloof in robuustheid. Iets dat in sommige gevallen werkt – zoals rijden op de snelweg –, is niet met evenredige stappen verwijderd van complexere gevallen – zoals rijden in de binnenstad. Metaforisch geformuleerd, je komt niet dichter bij de maan door in steeds langere bomen te klimmen.[454] Om bij de maan te komen,

451 Uit een interview met Andrew Ng in Ford 2018: 202.
452 Russell 2019: 151-152.
453 Russell 2019: 44.
454 Marcus en Davis 2019: 18-22, 66.

moet je andere methoden ontwikkelen. Het evenaren van de mens moet niet bij voorbaat uitgesloten worden, maar is met de huidige methoden nog ver buiten ons bereik. Fundamentele doorbraken zijn in de toekomst nodig om daar dichter bij te komen. Zoals in het tweede hoofdstuk is besproken, is alle hedendaagse AI nog steeds zogenoemde *narrow* of smalle AI, gericht op specifieke taken. Op verschillende van die taken streeft AI de mens al voorbij. Het is veel waarschijnlijker dat AI ons de komende jaren op allerlei smalle domeinen verder voorbij zal gaan dan dat AGI bereikt wordt.

Kernpunten – Beeld 'AI evenaart binnenkort de mens'

- Recente ontwikkelingen en doorbraken suggereren dat we dicht bij het evenaren van menselijke vermogens zijn, zogenoemd *artificial general intelligence* (AGI).
- Prominente wedstrijden en demonstraties verhullen echter de gecontroleerde omstandigheden waarin AI succesvol kan zijn.
- Er is een 'AI-kloof' tussen verwachting en realiteit. Die wordt gedreven door de projectie van hoe intelligentie bij mensen werkt op machines, door denkbeeldige vooruitgang wanneer we een mijlpaal verkeerd duiden en door het onterecht extrapoleren van simpele naar complexe vraagstukken.

'Kwaadaardige kunstmatige intelligentie kan zich tegen de mens keren'

Dit is misschien wel de grootste angst over AI die de samenleving heeft. Die wordt ook sterk gevoed door de verbeelding van AI in de populaire cultuur. Zoals we in het vorige hoofdstuk zagen, werd de term 'robot' voor het eerst gebruikt in een verhaal over mechanische arbeiders die zich tegen de mensheid keren. Al decennialang worden er films gemaakt met hetzelfde motief.

Deze verhalen berusten op een motief dat veel ouder is dan AI of computers. In het eerste hoofdstuk zagen we dat er ten minste vanaf de oude Grieken al verhalen bestaan over kunstmatige vormen van leven. In veel van die verhalen zit een dystopisch element. Het creëren van kunstmatig leven is vaak gezien als een overschrijding van grenzen waar straf op staat. Dat gebeurt dan ook in de verhalen van Prometheus, Daedalus en Medea. In die traditie staat ook het modernere verhaal van Frankenstein – met de ondertitel 'The Modern Prometheus' – van Mary Shelly. Dr. Frankenstein creëert een kunstmatig leven dat uiteindelijk zijn schepper doodt. De angst voor kwaadaardige AI lijkt dus voort te bouwen op een lange traditie van angstbeelden.

Ook een ander fenomeen, de zogenoemde *'uncanny valley'*, draagt bij aan deze angst. Dat fenomeen gaat over onze relatie tot machines die menselijke eigenschappen of gedrag vertonen. Een menselijke vorm creëert sympathie, maar wanneer een machine te dichtbij die vorm komt, verandert ze in angst en weerzin. Pas wanneer machines niet meer van mensen te onderscheiden zijn, kunnen we die 'griezelige vallei' weer verlaten. Ook dit fenomeen voedt de angst voor kwaadwillende AI.

Onderzoekers als Nick Bostrom en Max Tegmark wijden aan dergelijke scenario's verschillende gedachtenexperimenten.[455] Tegelijkertijd geven ze ook aan dat het speculatieve scenario's zijn. In films heeft de kwaadaardige AI vaak een mensachtige ('*humanoid*') vorm van een robot of een pratende computer. Alhoewel AI wel degelijk de vorm van een robot aan kan nemen, is een dergelijke belichaming voor de ontwikkeling van AI niet noodzakelijk. Zeer krachtige AI zal eerder de vorm van ontastbare algoritmes aannemen dan die van een lijfelijke machine tegenover ons.

Naast de vorm schrijft deze mythe over kwaadaardige AI ook allerlei andere menselijke eigenschappen aan AI toe, waarvoor er geen reden is om aan te nemen dat ze die zal ontwikkelen, zoals heerszucht, vrijheidsdrang, jaloezie en angst voor de dood. Steven Pinker formuleert dit bezwaar als volgt:

> "[H]et scenario dat robots superintelligent worden en mensen tot slaven zullen maken is even zinvol als de angst dat omdat vliegtuigen het vliegvermogen van adelaars hebben gepasseerd, zij op een dag uit de lucht zullen neerdalen om onze kuddedieren te grijpen. De … dwaling is een verwarring van intelligentie met motivatie, van inferenties met doelen, denken met willen. Zelfs als we supermenselijke intelligente robots zullen maken, waarom zouden zij dan hun meester *willen* overheersen en de wereld overnemen? Intelligentie is het vermogen om nieuwe middelen toe te passen om een doel te bereiken. Maar de doelen zijn extern aan de intelligentie: Slim zijn is niet hetzelfde als iets willen."[456]

Yann LeCun geeft aan dat de drang om de wereld te overheersen niet correleert met intelligentie, maar met testosteron.[457] Een gerelateerd bezwaar is dat de scenario's van kwaadwillende AI veronderstellen dat we het niveau van AGI bereikt hebben en, zoals we hiervoor hebben gezien, is dat nog niet in zicht.

---

455 Bostrom 2016; Tegmark 2017.
456 Marcus en Davis 2019: 30.
457 Uit een interview met Yann LeCun (Ford 2018: 135).

Een belangrijk bezwaar tegen deze mythe, juist omdat die zo krachtig is, is dat deze de aandacht op iets speculatiefs richt en afleidt van serieuzere dreigingen die momenteel al wel bestaan. Een machine hoeft namelijk geen kwade wil te hebben om ons leven wel degelijk te bedreigen. Een raket die op iemand gericht is, hoeft helemaal geen slechte intenties te hebben om iemand te kunnen doden. Het probleem is dan ook niet zozeer dat AI eigen, kwaadaardige doelen ontwikkelt, maar dat ze heel behendig is in het bereiken van doelen die mensen erin hebben gestopt en die gevaren opleveren of onvoldoende doordacht zijn.

Dit brengt ons op het probleem van *value alignment*, het laten samenvallen van de doelen in AI met onze eigen doelen. Het rigoureus navolgen van bepaalde doelen door AI kan namelijk andere doelen in gevaar brengen. Russell spreekt in dit verband van het 'koning Midas'-probleem. Koning Midas wenste dat alles wat hij aanraakte, in goud veranderde. Zijn doel van grote welvaart bereikte hij hiermee, maar toen ook eten en zijn familieleden in goud veranderden, kwam hij erachter dat dit ene doel met andere conflicteerde.[458]

Een steeds beter werkende AI die destructief wordt in het streven naar bepaalde voorgeprogrammeerde doelen, is dan ook een reëler risico dan kwaadwillende AI. Zoals Norbert Wiener stelde: "Menselijk onvermogen heeft ons beschermd van het volledige destructieve effect van menselijke dwaasheid".[459] Nu we het vermogen bezitten om machines op geavanceerde wijze doelen te laten realiseren, worden we geconfronteerd met onze slechtdoordachte doelstellingen. Een bekende voorstelling van dit probleem is Nick Bostroms gedachtenexperiment van de paperclipmachine. Hij stelt een zeer intelligente machine voor die als doel heeft zoveel mogelijk paperclips te produceren. Om dat doel te realiseren, zou het kunnen dat de machine vervolgens eerst de mensheid wegvaagt, zodat ze daarna in alle rust en zonder tegenwerking alle materie die ze vindt tot paperclips om kan vormen.[460] Rechtlijnige AI zonder eigen doelen is een groter gevaar dan AI met snode plannen.

Russell geeft een krachtig voorbeeld van de destructieve effecten van een simpel algoritme, namelijk om content te selecteren op sociale media. Het doel daarvan is om advertentiekosten te maximaliseren door het aantal *click-throughs* te vergroten. Dat lijkt nog relatief onschuldig als het algoritme daardoor aan de slag gaat om content te selecteren voor mensen die deze content het meest interessant zullen vinden. Dit algoritme bereikte het doel echter op een andere manier, namelijk door de voorkeuren van mensen zo te veranderen dat hun gedrag

458 Russell 2019: 137.
459 Wiener 1964.
460 Bostrom 2016.

voorspelbaarder werd. Omdat mensen met extremere politieke meningen beter voorspelbare voorkeuren hebben, zette het algoritme mensen er daarom toe aan zich voor extremere content te interesseren. Dit is een belangrijke factor bij het vijandige politieke klimaat op sociale-mediaplatformen. Het brengt risico's voor de democratie met zich mee, maar volgt zonder kwade intenties uit de optimalisering van een simpel doel, advertentie-inkomsten maximaliseren.[461]

Ook rondom de zelfrijdende auto krijgen acute vraagstukken te weinig aandacht in vergelijking met speculatieve scenario's. Veel van de discussies daarover gaan over het zogenoemde trolleyprobleem: Wat moet een zelfrijdende auto doen wanneer een ongeluk onvermijdelijk is en er moet worden gekozen wie er wordt gespaard? Ook hier is veel speculatie over wat de juiste waarden zijn, of die universeel zijn en of mensen auto's zouden kopen die de bestuurder opofferen om het leven van anderen te redden. Terwijl er interessante filosofische discussies over op te zetten zijn, zijn er veel andere meer acute uitdagingen rondom de zelfrijdende auto. Bovendien zijn er nu al simpeler vormen van bestuurdersassistentie op de markt die slachtoffers maken, waar onze aandacht dus beter naar uit zou kunnen gaan.[462]

Berichtgeving en de frames die worden gebruikt bij de communicatie over AI kunnen bij het bredere publiek het idee oproepen dat AI zich kwalijk ontwikkelt. Dat is een mythe. De programma's van Facebook waren niet tegen de mensheid aan het samenzweren, evenmin als autonome wapens de wereld over willen nemen. De handelingen van die wapens zijn levensgevaarlijk, maar zijn technisch gezien niet anders dan een schaakcomputer die het doel heeft meegekregen om te winnen, een zet uitrekent en die vervolgens uitvoert.

Kernpunten – Beeld 'AI kan kwaadaardig worden'

- Mede door populaire media is kwaadaardige AI een wijdverspreide angst onder de bevolking die diepe historische wortels heeft. Specifieke terminologie als 'killer robots' jaagt deze angst aan.
- Dit angstbeeld projecteert menselijke kenmerken en intenties op AI terwijl daarvoor weinig reden is.
- Kwaadaardige AI veronderstelt bovendien AGI waar we nog ver van verwijderd zijn.
- Zonder kwade intenties kan AI echter wel gevaarlijk zijn door verkeerde doelen te realiseren of door bepaalde doelen te realiseren ten koste van andere.

461 Russell 2019: 8-9.
462 Onderzoeksraad voor Veiligheid 2019

Dat het een mythe is dan AI zich kwalijk ontwikkelt, betekent echter niet dat we dit beeld niet serieus moeten nemen. Zoals de geschiedenis van systeemtechnologieën leert, zijn woorden, associaties en angstbeelden vaak zeer invloedrijk geweest. Soms hebben die ertoe geleid dat het publiek zich tegen een technologie keerde. Om uiteindelijk als samenleving de vruchten te kunnen plukken van een nieuwe systeemtechnologie als AI, is demystificatie dus van belang.

Generieke mythen over digitale technologie

'Technologie moet zo min mogelijk gereguleerd worden'

De vorige vijf beelden gingen specifiek over AI, drie over hoe ze werkt en twee over wat voor effecten ze in de toekomst heeft. AI is een grote nieuwe technologische ontwikkeling die onderdeel is van bredere digitale technologie. Leidende platformen van eerdere digitale technologieën zijn nu dan ook de leiders in AI. Vanwege die verwevenheid is het belangrijk om te kijken naar bredere beelden over (digitale) technologie die hun oorsprong vinden in Silicon Valley. Demystificatie daarvan zal ons ook helpen beter zicht te krijgen op AI.

Een eerste krachtig beeld over technologie, afkomstig uit Silicon Valley, is dat technologie zo min mogelijk gereguleerd moet worden. Dat idee kan op verschillende manieren beargumenteerd worden. Het kan volgen uit een benadering van techno-determinisme, het idee dat technologie een autonome werking heeft waaraan de samenleving zich simpelweg heeft aan te passen. Samenlevingen die dat niet doen en een technologie aan banden leggen, zullen achterblijven. Het motto van de wereldtentoonstelling in Chicago in 1933 was: 'Wetenschap vindt – Industrie past toe – de Mens past zich eraan aan'.[463] Weinig mensen zullen het tegenwoordig zo stellig formuleren, maar mildere varianten van techno-determinisme worden nog volop omarmd.

We zien ook een instrumentele benadering van technologie.[464] Daarbij is technologie juist niet bepalend voor de samenleving, maar wordt ze voorgesteld als een instrument waarvan de mens zelf bepaalt hoe het te gebruiken. Een hamer kan bijvoorbeeld gebruikt worden om een huis mee te bouwen of om iemand te doden; de keus is aan de mens.

In beide benaderingen, de techno-deterministische en de instrumentele, zit een kern van waarheid. Aan de beroemde techniekfilosoof Marshall McLuhan wordt vaak het volgende citaat toegeschreven: "*First we shape our tools and thereafter*

463 Zuboff 2019: 15.
464 Zie voor een bespreking van techno-determinisme en de instrumentele benadering: WRR 2011.

our tools shape us."[465] Wat hij in zijn werk suggereerde, is dat samenleving en techniek niet van elkaar te scheiden zijn. Ze zijn ten diepste met elkaar verweven. De geschiedenis van systeemtechnologieën leert ons ook dat de inbedding van een nieuwe technologie juist vraagt om een proces van wederzijdse aanpassing, waarvan het stellen van normen en regulering een noodzakelijk onderdeel is.

Alhoewel ze radicaal van elkaar verschillen, leiden zowel de techno-deterministische als de instrumentele benadering echter tot dezelfde conclusie, namelijk dat technologie zo min mogelijk gereguleerd moet worden. Omdat regulering futiel is in de eerste benadering en omdat deze zich op gebruik en niet op de technologie moet richten volgens de tweede benadering. De geschiedenis leert ons ook dat een ideologie van het zo vrij mogelijk laten van technologie bij elke systeemtechnologie opkomt en dat die benadering later correctie vereist.[466] Rondom de hedendaagse technologie lijkt die correctie nu langzaam plaats te vinden. In hoofdstuk 7 gaan we specifiek in op de opgave van regulering. Hier zullen we eerst laten zien waar de mythe dat er geen regels nodig zijn, vandaan komt en hoe dat frame rondom de hedendaagse technologie wordt gebruikt om een bepaalde agenda te legitimeren die mogelijk publieke waarden onder druk zet.

Jonathan Taplin heeft het gedachtengoed van Silicon Valley op mooie wijze gedocumenteerd. Hij beschrijft hoe Facebooks CEO Mark Zuckerberg in 2011 de Arabische lente aangrijpt voor een techno-deterministische redenering. Zuckerberg uitte lof voor de hulp bij het afzetten van dictators en contrasteerde die met de vrees voor het verzamelen en delen van informatie. "Je kunt niet sommige dingen die je goed vindt aan het internet isoleren en de dingen die je niet goed vindt controleren", aldus Zuckerberg.[467] De slogan van Google was aanvankelijk: '*don't be evil*'. Deze framing van techbedrijven als een kracht voor het maatschappelijk belang heeft als doel om zo vrij mogelijk te blijven van vormen van controle.[468]

Veel bedrijven uit Silicon Valley willen niet alleen zo min mogelijk regulering, maar verzetten zich ook tegen allerlei vormen van bestaande wetten en normen. Zo lanceerde Uber de app op allerlei plekken waar dit in strijd was met de taxiwetgeving. Het bedrijf ontwikkelde zelfs een programma genaamd Greyball

---

465 Deze uitspraak is overigens niet letterlijk terug te vinden in het werk van McLuhan; zie daarover het voorwoord van Lewis Lapham in McLuhan 1994 [1964].
466 Zo werd aanvankelijk ook over het internet gesproken (Stikker 2019).
467 Taplin 2017: 221.
468 Taplin 2017: 97.

om uit te rekenen hoe plotse controles door politiediensten het beste omzeild konden worden.[469]

De botsing met bestaande regels en conventies zit diep in de cultuur van Silicon Valley. Dit dogma uit zich in positieve termen als 'disruptie' en hangt samen met de hackersbeweging. De eerste brief van Zuckerberg aan investeerders toen het bedrijf naar de beurs ging, heette 'The Hacker Way'. Daarin gaf hij aan dat hacken onterecht een negatieve connotatie had. Het verstoren van de bestaande orde was zelfs een doel. Tot 2014 was het interne motto van Facebook '*move fast and break things*'. In een interview uit 2009 stelde Zuckerberg dat "als je niets breekt, je niet snel genoeg beweegt".[470]

Nog sterker komt het verzet tegen de bestaande orde tot uiting in het boek *From Zero to One* van PayPaloprichter Peter Thiel. Hierin beschrijft hij trots hoe vier van de zes mensen die dat bedrijf opstartten, op de middelbare school bommen hadden gemaakt.[471] Peter Thiel is nog steeds een belangrijke investeerder in Silicon Valley. De mensen met wie hij PayPal heeft opgericht, ook wel de 'PayPalmaffia' genoemd, zijn daarna belangrijke functies gaan vervullen bij allerlei bedrijven als Tesla, YouTube, Facebook en Palantir, een softwarebedrijf dat wereldwijd vooral actief is in het veiligheidsdomein.

Taplin toont aan dat deze mentaliteit volgt uit libertarische overtuigingen die een zo klein mogelijke overheid voorstaan en te traceren zijn tot de filosofie van Ayn Rand. Rand propageerde een zo vrij mogelijke markt geleid door de grootste ondernemers. Die ondernemers "vragen niet wie ze toestaat om iets te doen, maar eerder wie ze tegenhoudt". Toestemming vragen voor innovatie hoort er dus niet bij. Van Peter Thiel is bekend dat hij een aanhanger is van haar gedachtengoed.[472]

Een afkeer voor inmenging en regulering door de overheid is in veel ondernemingen van Silicon Valley zichtbaar. Het kenmerkt de *cypherpunk*-beweging die middels technologieën als cryptografie overheidsinmenging onmogelijk wil maken. Computerspecialist Ryan Lackey verhuisde in 1999 naar Sealand, een eiland voor de Engelse kust dat zich, zonder erkend te zijn, onafhankelijk heeft verklaard.[473] Van Google-oprichter Larry Page is bekend dat hij onderzoek heeft laten verrichten naar autonome stadstaten. Een recent voorbeeld van deze poging om buiten het bereik van overheden te komen is het Seasteading

469 Broussard 2019: 74.
470 Business Insider, 15 oktober 2010.
471 Taplin 2017: 76.
472 Taplin 2017: 227.
473 Rid 2016: 287.

Institute, dat ernaar streeft om buiten territoriale wateren een artificieel eiland te bouwen zonder overheid.

Als onderdeel van de *cypherpunk*-beweging publiceerde Timothy C. May in 1988 het *Crypto Anarchist Manifesto*. Daarin verwees hij naar het beeld van de oude Amerikaanse frontier als een vrij en wetteloos terrein. Een kleine uitvinding, prikkeldraad, maakte het echter mogelijk om toch grenzen en privébezit af te schermen. Volgens May is het internet de nieuwe frontier. De kleine uitvinding van cryptografie zou nu echter aan de kant van vrijheid staan, door grenzen en bezit online onmogelijk te maken.[474]

Diezelfde metafoor werd ook gebruikt in een tekst die de internetpionier Stuart Brand in 1990 publiceerde, genaamd *Crime and Puzzlement: In Advance of the Law on the Electronic Frontier*. Cyberspace werd hierin expliciet vergeleken met het Wilde Westen van negentiende-eeuws Amerika. De auteur John Perry Barlow richtte daarna de Electronic Frontier Foundation op en publiceerde later de *Declaration of Independence of Cyberspace*. In deze tekst zegt hij een vertegenwoordiger van de toekomst te zijn die overheden komt vertellen dat zij in cyberspace geen soevereiniteit hebben.[475]

Weer een ander voorbeeld waarin we de libertarische afkeer van regels herkennen, is in onlinepiraterij. Kim Dotcom, oprichter en eigenaar van de grote muziekpiraterijsite Megaupload totdat die werd neergehaald, schreef een rapnummer waarin hij zichzelf neerzet als een verdediger van vrije meningsuiting en zichzelf vergelijkt met Martin Luther King.[476]

In deze voorbeelden van overtuigingen in Silicon Valley zien we dus een spanning met zaken als privébezit, privacy en een sterke staat die bijvoorbeeld aan herverdeling van inkomen doet ten faveure van een libertarisch Wilde Westen.

In een aantal gevallen gaat deze ideologie nog verder tegen cruciale publieke waarden in. Hoewel dat natuurlijk niet voor heel Silicon Valley geldt, zijn er mensen zoals de genoemde invloedrijke Peter Thiel die vraagtekens plaatsen bij de democratie. Zo stelde hij dat hij "niet meer gelooft dat vrijheid en democratie verenigbaar zijn". Zijn persoonlijke voorkeur ligt duidelijk bij de eerste. In een tekst voor de website van het Cato Instituut, een economisch rechtse denktank, stelde hij zelfs dat sinds 1920 de groei van de welvaartsstaat en het vrouwenstemrecht de notie van 'kapitalistische democratie' iets tegenstrijdigs hebben

474 May 1994.
475 Rid 2016: 240-244.
476 Taplin 2017: 174.

gemaakt.[477] Op de Cato-website schreef hij dat het tegenwoordig de grote taak van libertariërs is om een ontspanning van politiek in alle vormen te vinden, een ontsnapping van het "niet-denkende demos" en dat er een dodelijke race gaande is tussen politiek en technologie.[478]

Uiteraard zijn dit extreme standpunten en zullen velen in Silicon Valley hem hierin niet volgen. Er zijn verschillende stromingen binnen Silicon Valley en veel mensen daar zijn inmiddels overtuigd van de noodzaak van overheidsingrijpen. Maar bovengenoemde standpunten zijn wel die van een invloedrijk figuur en in zwakkere vorm worden ze wel op allerlei plaatsen nog steeds uitgedragen, met name door grote technologiebedrijven. Samuel Freeman betoogt dat het recente libertarische gedachtengoed niet meer liberaal is. Het lijkt meer op een vorm van feodalisme die een gedeelde publieke ruimte wil vervangen door bilaterale individuele contracten tussen bedrijven en consumenten.[479]

Veel van bovengenoemde standpunten over het vrij laten van technologie komen specifiek terug in de context van AI. Ook hier wordt vaak betoogd dat regelgeving niet nodig, niet mogelijk is of schadelijk is en een samenleving op achterstand brengt. Op wat voor manier dat problematisch is, bespreken we in hoofdstuk 7. Hier is het belangrijk om ons te realiseren dat, net als bij eerdere technologieën, AI ook gepaard gaat met een expliciete ideologie die elke vorm van regelgeving afwijst en op gespannen voet kan staan met de democratie. De geschiedenis leert dat dat tot allerlei gevaren en ongelukken leidt. Bovendien staan regels en normen niet haaks op de ontwikkeling van een technologie, maar kunnen ze het gebruik ervan juist faciliteren. Om tot geschikte regulering te komen is het behulpzaam om ons bewust te zijn van de bronnen, de werking en de gevaren van de mythe waarin het nut daarvan wordt weersproken.

Kernpunten – Beeld 'Technologie moet zo min mogelijk gereguleerd worden'

- Vanuit zowel techno-deterministische als instrumentele benaderingen van technologie wordt de suggestie gewekt dat technologie zo min mogelijk gereguleerd moet worden.
- De cultuur van disruptie, hacking en libertarische overtuigingen zet Silicon Valley vaak op gespannen voet met de bestaande maatschappelijke en politieke orde.
- Er zijn zelfs stromingen en ontwikkelingen in Silicon Valley die slecht met democratische controle te verenigen zijn.

477 Taplin 2017: 70.
478 Thiel, 13 april 2009.
479 Freeman 2001.

'There is no alternative' (TINA)

Nauw verbonden met het vorige beeld is een tweede algemeen beeld over de aard van technologie. Een van de beschreven bronnen voor het niet reguleren van technologie is het techno-determinisme, dat stelt dat de samenleving zich aan de technologie moet aanpassen. Gerelateerd daaraan is het idee dat de vorm en de uitwerking die technologie tegenwoordig heeft, inherent zijn aan de technologie zelf en dat er daarmee dus geen alternatieven zijn. Denk aan het bestaan van zeer grote bedrijven, het verzamelen van data, advertenties als een bron van inkomsten en markten als de bron van alle innovatie; het zijn allemaal aspecten van de invloed van hedendaagse technologie op de samenleving die inherent zouden zijn aan de technologie. Als we de voordelen ervan willen plukken, dan zullen we al die aspecten ook moeten accepteren, aldus dit beeld.

Evgeny Morozov onderscheidt de fysieke infrastructuur van het internet van wat hij het mythische internet noemt. Dit laatste is een complexe vergaarbak waar allerlei wensen en projecties aankleven, maar die volgens hem vrij weinig te maken heeft met het fysieke internet. 'Het internet', zoals de mythische variant vaak wordt aangeduid, heeft geen duidelijke betekenis en kan vrijwel alles omvatten wat er online gebeurt, van businessmodellen tot de strijd om netneutraliteit en allerlei met internet verbonden technologieën.[480] 'Het internet' is een retorische constructie, een mythe, die ervoor dient om een goed begrip en een kritische blik onmogelijk te maken.

Natuurlijk kunnen vanuit dit beeld dat er geen alternatief is, wel degelijk allerlei variaties binnen technologie bestaan. Terwijl het businessmodel van Google en Facebook bijvoorbeeld om advertenties draait, geldt dat niet voor een bedrijf als Apple. Tussen sociale-mediaplatformen bestaan ook allerlei verschillen. Maar dit beeld stelt wel dat de fundamentele organisatie van hedendaagse technologie onveranderlijk is en dat alternatieve modellen daarvoor, zoals het niet verzamelen van data of gebruikers zelf daar de eigenaar van laten zijn, onrealistische ideeën zijn.

Het gaat er in dit geval niet om of specifieke alternatieven realistisch zijn. Wel nemen wij de gedachte dat de huidige verschijningsvorm van technologie noodzakelijk is kritisch onder de loep en laten we zien dat dat een mythe is. In *De publieke kern van het internet* heeft de WRR de kernbestanddelen en diepere lagen van het internet onderscheiden van de bovenbouw waar grote

480 Morozov 2013: 17-18.

technologiebedrijven gebruik van maken.[481] Marleen Stikker stelt regelmatig de keuzes die ten grondslag liggen aan het bestaande internet, ter discussie.[482]

Een aantal auteurs plaatst de laatste jaren vraagtekens bij het idee van een noodzakelijke verbinding van technologie en de vrije markt, en ideeën over private bedrijven als de belangrijkste bron van innovatie. Mariana Mazzucato betoogt bijvoorbeeld dat veel hedendaagse innovatie niet oorspronkelijk uit de markt komt, maar vanuit de overheid. Terwijl markten er goed in zijn om innovatie te vercommercialiseren, ligt aan innovatie een lang proces van fundamenteel onderzoek ten grondslag dat voor de markt te risicovol is en te veel gericht op de lange termijn. Overheidsfinanciering is daarvoor nodig. De overheid staat aan de basis van duurzame energie, zoals zonnepanelen, maar ook van veel innovatie in de biotechnologie en de nanotechnologie. Naast studies van deze domeinen laat ze zien hoe allerlei cruciale onderdelen van de iPhone ontstaan zijn uit door de overheid gefinancierd onderzoek: dat geldt voor het internet, de touchscreen en GPS en zelfs de spraakassistent Siri, die afkomstig is van het wetenschappelijke lab SRI.[483] Zo prikt Mazzucato de mythe door dat alleen grote technologiebedrijven alle innovatie die er nu is, kunnen ontwikkelen. Daarbovenop stelt zij de vraag of het terecht is dat de overheid – en dus het publiek – de risico's draagt voor fundamentele innovaties en private ondernemingen zich daarvan alle winst toe-eigenen.

Ook Shoshana Zuboff laat zien dat er beslissingen ten grondslag liggen aan de vorm van de hedendaagse technologie en dat er daarmee dus alternatieve vormen van inrichting zijn. Ze beschrijft een project van Georgia Tech uit 2000, getiteld 'Aware Home'. Het was een vroege vorm van allerlei *smart home*-technologie die nu bestaat, zoals zogenoemde slimme thermostaten en virtuele assistenten. Zuboff merkt op dat het project echter een volledig ander model volgde en dat de data bijvoorbeeld volledig eigendom bleven van de bewoners.[484] In haar brede studie laat Zuboff zien hoe gaandeweg technologie verweven is geraakt met andere maatschappelijke ontwikkelingen, waardoor deze haar huidige vorm heeft aangenomen. Ze beschrijft hoe eerst de neoliberale marktekonomie ermee verbonden raakte. Na 11 september 2001 zetten overheden vervolgens in op het verzamelen van data en het surveilleren van de bevolking, waardoor zij gelieerd raakten aan bedrijven in Silicon Valley die daar goed in zijn. Zowel de neoliberale markt als de dataverzameling voor surveillance zijn externe ontwikkelingen die niet inherent zijn aan het functioneren van

481 WRR 2015.
482 Stikker 2019.
483 Mazzucato 2014.
484 Zuboff 2019: 5-6.

technologie zelf. Haar kritiek richt zich dan ook niet zozeer op de technologie als op de bezitters van technologieën en de keuzes die zij maken – in haar woorden de "*puppet masters, not the puppet*".[485]

Tekstbox 4.2 – Acceleratie

Een ander soort bron die ook teruggrijpt op het historisch meer publieke karakter van innovatie, is het *Accelerationistisch Manifesto*, een tekst van Alex Williams en Nick Srnicek uit 2013. De tekst is utopisch over de mogelijkheden van technologie om allerlei vraagstukken op te lossen, een uitgangspunt dat wij als volgende beeld kritisch zullen behandelen.

Relevant hier is dat de auteurs de aandacht richten op de discrepantie tussen de grote beloften rondom technologie nu en het feitelijke gebruik ervan voor het maken van onnodige gadgets en het genereren van advertentie-inkomsten. Dat wijten zij aan de koppeling van technologie aan de neoliberale ideologie. De auteurs pleiten ervoor om inspiratie te halen uit eerdere perioden, zoals de jaren zestig toen de doelen waarvoor technologie werd ingezet, zoals een mens naar de maan brengen, bredere maatschappelijke belangen meenamen.

Zij sluiten daarmee ook aan bij het werk van Mazzucato die voorstelt niet alleen te kijken naar het tempo van innovatie, maar ook naar de richting ervan. Zij spreekt eveneens van '*man on the moon projects*' als model voor de inzet van technologie voor publieke doelen.

Hoe kunnen we vanuit deze brede blik op technologie naar AI specifiek kijken? De historisch grote rol van de overheid bij de ontwikkeling van technologie is zeker ook bij AI waar te nemen. In het bijzonder het Amerikaanse leger en de onderzoekstak daarvan, DARPA, speelde een cruciale rol in de ontwikkeling van AI.[486] De organisatie van AI in China, Japan en Zuid-Korea toont bovendien aan dat ook tegenwoordig die technologie veel sterker vanuit de overheid aangestuurd kan worden. Of dat wenselijk is, is een andere vraag. Relevant is hier dat de exclusieve koppeling aan grote private ondernemingen niet het enige mogelijke model is. De geschiedenis van AI leert bovendien dat naast de publieke oriëntatie er ook nog een ander model rondom deze technologie heeft bestaan.

485 Zuboff 2019: 14.
486 Weinberger 2019.

Een van de bedenkers daarvan was Douglas Engelbart. In 1968 gaf hij een 100 minuten durende presentatie die bekend is komen te staan als '*the mother of all demos*'. Decennia geleden stelde hij allerlei nieuwe technologieën voor, zoals de muis, windows, videoconferencing en hypertext. Vooral belangrijk aan deze presentatie is dat hij hierin computertechnologie in een andere context plaatste. Het oude mainframe was een apparaat dat werd toegepast in grote overheidsorganisaties en gecentraliseerde organisaties als IBM. Engelbart legde de visie neer van de personal computer als een apparaat voor gebruik door individuele burgers en dus met een decentraliserende werking.[487]

Achter de camera van de demo stond Stewart Brand, oprichter van het blad de *Whole Earth Catalog* en inspirator voor de eerste generatie internetpioniers. Dat blad speelde een centrale rol bij het overplaatsen van het idealisme van de hippiebeweging naar de computertechnologie.[488] Zo werd de computer, die daarvoor onderdeel was van 'Koude Oorlog-technocratie', nu een onderdeel van een verlangen naar persoonlijke ontplooiing, samenwerking en gemeenschap.[489] Cyberspace was niet langer alleen relevant voor militaire projecten en ruimtevaart, maar nu ook voor de *counterculture* van San Francisco.[490]

Jonathan Taplin stelt dat libertariërs voorbijgaan aan het feit dat het internet initieel is bedacht en gefinancierd door de overheid en vervolgens is overgenomen door idealisten en academici die geen interesse hadden in winst. Vroege ontwikkelaars zoals Tim Berners-Lee, een van de vaders van het internet, schreven gratis code uit idealisme.[491] Nog regelmatig uit Berners-Lee zich kritisch over de huidige vorm die het internet heeft aangenomen en zet hij zich in voor alternatieven.[492]

Gaandeweg is digitale technologie veel meer verbonden geraakt met een meer libertarisch en technocratisch marktmodel. Het contrast tussen verschillende visies komt mooi naar voren in een gesprek tussen Douglas Engelbart en de beroemde AI-pionier Marvin Minsky. Die laatste stelde dat hij van plan was om machines intelligent en bewust te maken. Engelbart antwoordde hierop: "Je gaat al die dingen doen voor machines? En wat ga je doen voor mensen?"[493]

---

487 Rid 2016: 173.
488 In 1995 legde Brand de connectie tussen counterculture en de personal computer in een essay voor *Time Magazine* getiteld *'We owe it all to the hippies'* (Brand, 1 maart 1995).
489 Turner 2006.
490 Rid 2016: 166.
491 Taplin 2017: 54.
492 Hern, 12 maart 2019.
493 Taplin 2017: 56.

Wat het voorgaande duidelijk maakt, is dat moderne technologie niet noodzakelijkerwijs verbonden is aan allerlei hedendaagse verschijningsvormen ervan. Het heeft ook in ten minste twee andere vormen bestaan. Veel van hoe technologie nu ingericht is, maakt er dus geen essentieel onderdeel van uit. Er zijn allerlei initiatieven om technologieën als AI voor andere doelen en in andere contexten in te zetten, en andere keuzes te maken over de inrichting ervan. In hoofdstuk 6 laten we zien hoe allerlei activisten zich inzetten om AI diverser en democratischer te maken.

Kernpunten – Beeld 'Er zijn geen alternatieven'

- De mythe van 'het internet' stelt de technologie gelijk met de volledige hedendaagse inrichting van het internet.
- Technisch zijn er allerlei alternatieven mogelijk. Verschillende denkers laten zien hoe exogene factoren het internet nu vormgeven en er daarmee dus van te scheiden zijn.
- Ook historisch heeft digitale technologie andere vormen gekend, zoals het internet ten tijde van de Koude Oorlog en onder invloed van het idealisme van de jaren zestig.
- Er is een brede roep om tot een andere inrichting van het internet te komen.

'Technologie is de oplossing voor alle maatschappelijke vraagstukken'

Het laatste beeld van AI dat wij hier bespreken, is de overtuiging dat de oplossing voor grote en lastige maatschappelijke vraagstukken altijd in technologie te vinden is. Het is duidelijk dat technologie aan veel vraagstukken een belangrijke bijdrage kan leveren, en ook al doet. Tegelijkertijd is op veel plaatsen het geloof daarin doorgeschoten, met soms problematische gevolgen.

Een aantal auteurs heeft dit beeld op verschillende manieren gekarakteriseerd. Meredith Broussard hanteert de term 'techno-chauvinisme', de overtuiging dat technologie altijd de oplossing voor een probleem is.[494] Evgeny Morozov spreekt van 'technologisch solutionisme'. Hij opent zijn boek hierover met een citaat van Googles voormalige CEO Eric Schmidt: "In de toekomst zullen mensen minder tijd doorbrengen om technologie te laten werken (…) omdat het naadloos zal functioneren. Het zal er gewoon zijn. Het Web zal alles worden en het zal ook niets zijn. Het zal als elektriciteit zijn (…) Als we dit goed krijgen,

494 Broussard 2019: 7-8.

geloof ik dat we alle problemen van de wereld kunnen oplossen".[495] Een van de problemen die Morozov hierin onderscheidt, is dat solutionisme ervan uitgaat dat allerlei zaken problemen zijn, zonder dat dat altijd het geval hoeft te zijn. Een solutionistische benadering ziet overal in de werkelijkheid allerlei vormen van inefficiëntie, ambiguïteit en ondoorzichtigheid. Vaak is ondoorzichtigheid echter een voorwaarde voor privacy, beroepsgeheim of andere zaken waar we aan hechten. Gebrek aan efficiëntie biedt ruimte om te experimenteren en te oefenen, wat essentieel is voor veel menselijke activiteiten, zoals koken of een taal leren. De wil om zaken te veranderen herformuleert volgens Morozov allerlei complexe sociale situaties tot duidelijk gedefinieerde problemen, met uiteindelijk te berekenen oplossingen, of tot transparante en evidente processen die gemakkelijk geoptimaliseerd kunnen worden. De manier waarop technologie werkt, wordt daarmee een model voor andere domeinen: Wikipedia als een model voor het bedrijven van politiek, Facebook als een model voor burgerschap en Google als het model voor alle innovatie.[496]

Een gevaar van deze herformulering van allerlei maatschappelijke domeinen als technisch op te lossen problemen is dat hiermee belangrijke functies van die maatschappelijke domeinen in gevaar komen, onder meer doordat deze enorm versmald worden. Het doen van werk draait bijvoorbeeld niet alleen om de output. Wanneer een algoritme daarop stuurt, kan de efficiëntie toenemen, terwijl tegelijkertijd het werkplezier, een andere belangrijke functie van werk, ondermijnd wordt. We noemden eerder al het voorbeeld van de zorg, die niet alleen draait om het genezen van mensen, maar vaak ook een bron is van geruststelling of zelfs menselijk contact. Een algoritme zou het eerste kunnen optimaliseren, maar de andere functies juist kunnen doen verdwijnen als het menselijk contact overbodig maakt. Illustratief is ook de inzet van algoritmes binnen de rechtspraak. Dat de afhandeling van bepaalde zaken met behulp van AI kan worden geoptimaliseerd, zoals bij procedures over de administratiefrechtelijke afhandeling van verkeersboetes, wil niet zeggen dat het menselijk contact hier kan verdwijnen, zelfs niet bij dergelijke eenvoudige zaken.[497] Een ander voorbeeld zijn allerlei technologieën die door middel van surveillance en risicoinschattingen veiligheid en sociale harmonie in de samenleving stimuleren. Terwijl de criminaliteit hiermee mogelijk vermindert, zou de prijs van continue monitoring allerlei vormen van persoonlijk lijden kunnen zijn.

---

495 Morozov 2013: 1.
496 Morozov 2013: 5-6, 15.
497 Zie de diverse bijdragen in *Algoritmes in de Rechtspraak. Wat artificiële intelligentie kan betekenen voor de rechtspraak* (Raad voor de Rechtspraak 2019).

Tekstbox 4.3 – Corona-appathon
In reactie op de coronacrisis organiseerde de Nederlandse regering een appathon voor de ontwikkeling van een app. Hier dreigde ook een vorm van solutionisme waarbij meteen gegrepen werd naar nieuwe technologie, terwijl het maar de vraag is of die het meeste op zou leveren. Alternatieve benaderingen zouden beter kunnen aansluiten bij de behoeften van de GGD's. In Nieuw-Zeeland is bijvoorbeeld voor een lowtechbeleid gekozen, door alle burgers te vragen een dagboek van hun contacten bij te houden. Achteraf blijkt de app weinig bijgedragen te hebben aan de bestrijding van de pandemie. In een position paper voor de Tweede Kamer van april 2020 waarschuwde de WRR voor techno-solutionisme bij de ontwikkeling van een corona-app.

Broussard benadrukt het gevaar dat we iemands succes in een domein extrapoleren naar andere domeinen. Prominente technologiepioniers worden vaak als iconen behandeld die over van alles een mening kunnen hebben. Succes bij een wiskundige doorbraak of het ontwikkelen van een businessmodel betekent echter niet dat iemand expertise heeft op maatschappelijke vraagstukken of beleidsdomeinen. Sterker nog, de radicale toewijding aan bijvoorbeeld wiskundige oplossingen kan desastreus zijn voor (het vermogen tot) interactie tussen mensen.[498] Een laatste risico van techno-chauvinisme of solutionisme is dat dit de nadruk legt op revolutionaire plannen om zaken in de toekomst te bouwen en daarmee voorbijgaat aan het onderhouden en verbeteren van wat er al bestaat.

Kernpunten – Beeld 'Technologie als universele oplossing'
- Termen als techno-chauvinisme en -solutionisme duiden op de overtuiging dat alle weerbarstige maatschappelijke vraagstukken met technologie opgelost kunnen worden.
- Vanuit deze overtuiging worden allerlei zaken versimpeld of eendimensionaal uitgelegd, waardoor andere aspecten van de maatschappelijke orde onderbelicht of verstoord kunnen raken.
- Simpele kwantitatieve benaderingen of een nadruk op de nieuwste technologie haalt de aandacht weg van doeltreffender en niet-technologische benaderingen die soms veel beter kunnen werken.

498 Broussard 2019: 80.

4.3 Tot slot

Mythen zijn er altijd geweest en zullen er altijd blijven. Dat geldt ook in het geval van kunstmatige vormen van intelligentie. Het is dus onmogelijk om alle mythen over AI blijvend de wereld uit te helpen. Bovendien bestaat er niet zoiets als 'het reële beeld' over AI – daarvoor is de werkelijkheid te complex en onzeker. Dat betekent niet dat we niks kunnen doen tegen mythevorming. Het is namelijk wel mogelijk om onrealistische beelden te ontmantelen. De opgave van demystificatie bestaat er dan ook uit bij te dragen aan een beter begrip van wat deze technologie is en kan. Demystificatie levert daarmee een belangrijke bijdrage aan de overige opgaven: een beter bewustzijn van waar we het over hebben als het om AI gaat, helpt om na te denken over hoe we er verder mee omgaan. En vooral ook *dat* we ermee om blijven gaan. Onrealistische voorstellingen van AI kunnen immers leiden tot een algehele afkeer van de technologie, waardoor we reële kansen zouden missen. Aan de andere kant kan beperkt begrip ervoor zorgen dat er juist te grote risico's worden genomen en er slachtoffers vallen die we redelijkerwijs hadden kunnen voorkomen. Kortom, er is veel aan gelegen om AI te demystificeren.

In dit hoofdstuk zijn we ingegaan op de beeldvorming rondom AI en het belang van demystificatie. We hebben gezien hoe onrealistische beelden kunnen ontstaan doordat het generieke en nieuwe karakter van AI onze verbeelding ruim baan geeft, doordat AI soms wordt verbonden aan bestaande bronnen van argwaan, hoe indrukwekkende demonstraties kunnen bijdragen aan te hoge verwachtingen, en hoe het gebruik van bepaalde termen en frames bepalend kan zijn voor hoe we denken over AI. Daarnaast hebben we zeven beelden rondom AI behandeld. Drie gingen over de werking ervan (dat de technologie neutraal is, rationeel en een black box). Twee gingen over toekomstige verwachtingen (het evenaren van menselijke intelligentie en het gevaar van kwaadaardige AI) en drie waren bredere beelden over technologie die vaak in de context van AI terugkomen – dat ze ongereguleerd dient te zijn, maar een enkele vorm kan aannemen en de oplossing is voor alle maatschappelijke problemen. De beelden zijn erg divers. Terwijl sommige makkelijk door te prikken zijn, geldt dat niet voor andere, met name voor beelden vanuit de ideologie van Silicon Valley.

Ten slotte hebben we gezien dat verschillende actoren een rol spelen bij deze opgave. Deels ligt die rol bij de samenleving zelf, door grotere ervaring en gewenning met de technologie op te doen. Wetenschappers, scholen en media vervullen hier in het bijzonder een belangrijke rol bij. Ook de overheid kan bijdragen aan demystificatie door te investeren in kennis en publieke campagnes, door instituten op te richten of de partijen die een belangrijke rol kunnen spelen te steunen. Wat deze opgave voor Nederland betekent en op dit moment van de overheid vraagt, behandelen we in deel 3. De overheid kan bijvoorbeeld op verschillende manieren bijdragen aan de ontwikkeling van kennis en

vertrouwdheid ten aanzien van AI, en heeft daarbij bovendien meerdere doelgroepen te bedienen. In het laatste hoofdstuk geven we de invulling van deze taak meer richting door actuele aandachtspunten te identificeren en doen we specifieke aanbevelingen om daarmee aan de slag te gaan.

5. Contextualisering

Contextualisering gaat over het gebruik van AI. Deze opgave is van groot belang bij de overgang van het lab naar de samenleving en betreft een complex en vaak onderschat proces. Contextualisering houdt in dat AI moet gaan functioneren in specifieke contexten, die allemaal hun eigen systemen, praktijken, regels en logica's kennen. Deze opgave van aanpassing en inpassing kost tijd, waardoor het vaak langer duurt dan we zouden denken voordat een nieuwe systeemtechnologie ook daadwerkelijk naar behoren werkt en haar weg vindt in ons dagelijks leven. De vraag van de contextualiseringsopgave is dus: *Hoe gaat AI werken?*

Bij het behandelen van deze vraag bespreken we allerlei toepassingen van AI. Zo zullen toepassingen in de zorg veel voorbijkomen. Een specifieke casus waarmee we het vraagstuk van contextualisering zullen illustreren, is een AI-toepassing waar hoge verwachtingen van zijn: de zelfrijdende auto (zie tekstbox 5.1). De ontwikkeling daarvan hangt namelijk minder af van mechanische aspecten van de auto, zoals de motor, en meer van intelligente algoritmes die beslissingen over routes nemen en dynamisch op de omgeving inspelen. Het voorbeeld van de zelfrijdende auto keert in verschillende tekstboxen steeds terug, om de centrale dimensies van de opgave van contextualisering te illustreren.

Wat betekent deze opgave nu als het specifiek over AI gaat? De auteur Kai-Fu Lee maakt een analogie met het eerdere contextualiseringsproces rondom elektriciteit. Na de ontdekkingen van Edison gingen tal van ondernemers aan de slag om allerlei industrieën revolutionair te veranderen. Ze pasten elektriciteit toe op het koken van eten, het verlichten van kamers en het aandrijven van industrieel gereedschap. In Kai-Fu Lee's analyse vereiste elektrificatie – het toepassen en laten werken van elektriciteit – vier 'inputs': fossiele brandstoffen om elektriciteit op te wekken, ondernemers om de technologie zakelijk te ontwikkelen, ingenieurs om de technologie te hanteren en een overheid om voor de onderliggende infrastructuur te zorgen. Naar analogie vergt AI volgens Kai-Fu Lee dus: data als grondstof, slimme ondernemers, AI-wetenschappers en een stimulerend overheidsbeleid.[499]

499 Lee 2018: 13-4.

Tekstbox 5.1 – Wat is een zelfrijdende auto?

Niveaus van autonomie

Wat een zelfrijdende auto is, is niet gemakkelijk af te bakenen. Een veel gehanteerd instrument daarvoor is het SAE-model, dat zes niveaus (nul tot en met vijf) van automatisering van auto's onderscheidt. Op niveau nul tot en met twee blijft de mens de bestuurder van de auto, tot op zekere hoogte ondersteund door software. Vanaf niveau drie rijdt de auto onder bepaalde omstandigheden zelf; de stap van twee naar drie is daarmee een grote sprong. Pas op level vijf is de auto in alle omstandigheden volledig geautomatiseerd. Dit laatste is nog verre toekomstmuziek. Op dit moment zijn de meeste moderne auto's niveau één, met zogenoemde *advanced driving assistance systems* (ADAS). Dit soort systemen kunnen automatisch de rijbaan aanhouden, helpen met inparkeren en hebben cruise control. Volgens een schatting van het ministerie van Infrastructuur en Waterstaat (IenW) uit 2018 had destijds 1 procent van de auto's een automatiseringsgraad van niveau twee met *adaptive cruise control* (ACC), waarbij de snelheid wordt aangepast aan de auto ervoor. Er zijn op dit moment nog geen commercieel verkrijgbare auto's met niveau drie.

Hedendaagse toepassingen

Een vorm van automatisering in het verkeer die al redelijk ver is, is 'truck platooning'. Een aantal vrachtwagens rijdt dan op korte volgafstand automatisch achter een eerste truck aan, die nog wel bemand wordt door een menselijke bestuurder. In Nederland worden sinds 2016 praktijkproeven gedaan met truck platooning op de openbare weg. In het kader van een Europees project werd in 2018 op initiatief van IenW 's wereld eerste test gedaan met truck platoons op de openbare weg over landsgrenzen heen. Technisch is platooning mogelijk, maar regelgeving verhindert op dit moment nog dat er voertuigen zonder bestuurder de weg op mogen.

Daarnaast is er een aantal voorbeelden van automatische robottaxi's. Die mogen alleen binnen een zeer beperkt gebied rijden ('geofencing') waarbinnen zij dus veel kilometers rijden en data verzamelen. Hieraan werkt onder andere het bedrijf Waymo, een dochteronderneming van Alphabet (Google), in de omgeving van de Amerikaanse stad Phoenix. Een bestuurder is daarbij wel voor de zekerheid aanwezig en dat bleek ook nodig. Een taxi stopte aan de verkeerde kant van de weg en had moeite om een terrein op te rijden waar veel activiteit was.

In Nederland was er een plan voor een hooggeautomatiseerde bus, de WEpod. Deze pendelbus moest gaan rijden in de regio Ede-Wageningen. Omdat de provincie de veiligheid niet kon garanderen, werd hiervoor echter geen vergunning afgegeven.

Een socio-technische benadering

Wij delen de elementen van contextualisering iets breder in dan Kai-Fu Lee. Hier gaan we uit van een *socio-technische ecosysteembenadering*. Deze valt uiteen in enerzijds een technisch ecosysteem van ondersteunende en emergente technologieën en anderzijds een sociaal ecosysteem van de menselijke context, dat we op macroniveau (werkgelegenheid en de productiviteitsparadox) en op microniveau (gedragsmatige context) kunnen beschouwen. Contextualisering is de opgave om een nieuwe technologie in deze verschillende ecosystemen in te bedden.

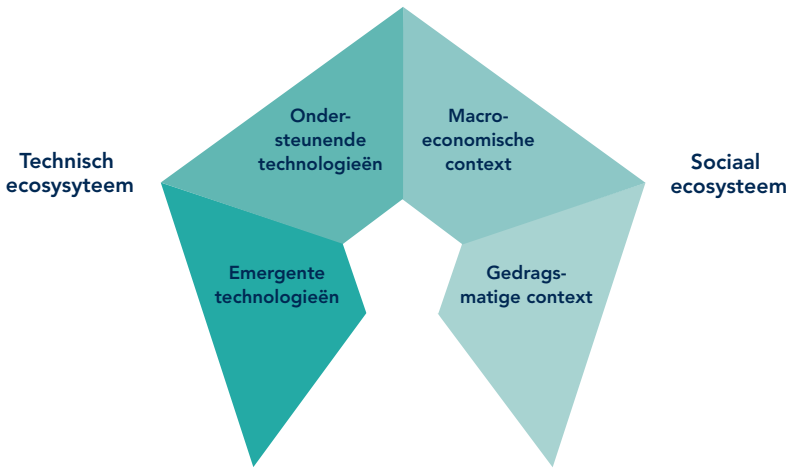
De waarde van een socio-technische benadering van AI wordt duidelijk als we deze met andere benaderingen contrasteren. Zo behandelt een streng afgebakende benadering van AI slechts de specifieke algoritmes die daartoe gerekend kunnen worden, zodat AI van bijvoorbeeld vraagstukken over data wordt gescheiden. Vanuit theoretische overwegingen valt dat te rechtvaardigen, maar zo'n strenge afbakening levert blinde vlekken op als het gaat over wat de technologie in de praktijk tot een succes maakt. Een ondersteunende technologie als data is daar namelijk noodzakelijk voor, ook al hoort data strikt genomen niet bij AI zelf. Als we een compleet beeld van de contextualiseringsopgave willen hebben, moeten we het bredere technische ecosysteem in beschouwing nemen.

Een tweede contrast kunnen we maken met de instrumentele benadering van AI die vaak doorklinkt in ethische analyses. Wanneer AI puur als instrument wordt beschouwd, wordt ze als neutraal opgevat; mensen kunnen de technologie kan dan namelijk goed of slecht gebruiken. Dat beperkt mogelijke antwoorden tot het opstellen van principes of regels voor goed gebruik. Het risico van deze benadering is dat de context waarbinnen de technologie werkt, buiten beschouwing wordt gelaten. Een ecosysteembenadering vraagt juist aandacht voor het feit dat hele omgevingen worden getransformeerd. Neem het internet. Een instrumentele nadruk op goed gebruik richt zich op ethische principes en formuleert bijvoorbeeld etiquetteregels voor onlinegedrag. Het internet heeft echter ook de publieke ruimte getransformeerd en de interactie tussen mensen beïnvloed. Een benadering die zich puur op de ethiek van goed gebruik richt, mist die bredere systematische veranderingen.

Een laatste andere benadering die aan het vraagstuk van contextualisering raakt, is het onderzoek naar 'AI readiness'. Oxford Insights publiceert daarover

jaarlijks een index om weer te geven hoe voorbereid landen zijn op AI. Die index raakt aan deze opgave, maar daarin zijn de niet-technische dimensies, ofwel het sociale ecosysteem, niet opgenomen. Binnen de technische dimensies behandelt deze index bovendien maar een beperkt aantal voorwaarden.[500] Een te strenge, instrumentele of te nauwe technische benadering laat dus belangrijke factoren buiten beschouwing die invloed hebben op de vraag wanneer AI kan werken. Wij benaderen die vraag daarom vanuit het socio-technische ecosysteem waarin AI gaat functioneren (zie figuur 5.1). Hieronder bespreken we eerst het technische ecosysteem (paragraaf 5.1) van AI, dat, zoals hiervoor aangegeven, uit twee dimensies bestaat. Daarna behandelen we het sociale ecosysteem (paragraaf 5.2).

Figuur 5.1 Het technische en sociale ecosysteem van AI



5.1 Het technische ecosysteem

Ondersteunende technologie

De eerste dimensie van het technische ecosysteem is ondersteunende technologie.[501] Deze ondersteunende technologieën behoren strikt genomen niet tot de nieuwe systeemtechnologie zelf, maar zijn wel een noodzakelijke voorwaarde om de systeemtechnologie te laten functioneren. Een hieraan verwant concept is dat van *enveloping*, waarbij een nieuwe technologie gaat functioneren door de

500 Er wordt wel gekeken naar het netwerk, maar niet specifiek naar data en hardware bijvoorbeeld.
501 Wij hanteren hierbij een breed begrip van technologie waaronder ook technisch ontwikkelde grondstoffen vallen.

omgeving op die technologie aan te passen, en niet per se door de technologie zelf te verbeteren.

AI in strikte zin gaat over de ontwikkeling van 'intelligente' algoritmes, zoals in deel 1 aan de orde kwam. Wat zijn dan andere technologieën waar AI op leunt? Enerzijds zijn dat de data die de grondstof voor AI vormen en anderzijds is dat de hardware die AI vereist.[502]

Bij hardware gaat het in eerste instantie om goede digitale netwerken. Dit betekent de aanwezigheid van een snel en betrouwbaar netwerk. AI bestaat uit complexe berekeningen die in veel gevallen zeer snel plaats moeten vinden. Denk aan beslissingen in het verkeer die in milliseconden plaats moeten vinden of de werking van machines in fabrieken. Dat alles moet niet alleen snel gebeuren, maar ook zonder hapering of 'latentie' (vertraging). Een netwerk dat even uitvalt op de weg, kan fataal zijn. Niet alle gebieden en plaatsen hebben een even goede dekking, zoals sommige dunbevolkte gebieden of omgevingen waar zware constructies of Faradaykooien signalen blokkeren. Voordat het mogelijk is AI ergens te implementeren, moet de vraag gesteld worden of het digitale netwerk aan de nodige vereisten daarvoor voldoet. Over het algemeen is het netwerk in Nederland van zeer hoog niveau, maar de aanwezigheid van betrouwbare netwerken vraagt continue aandacht.

Naast het netwerk gaat hardware ook over rekenkracht. Die komt van chiptechnologie en supercomputers. Van groot belang zijn de chips die ontwikkeld worden in de halfgeleiderindustrie. Met zulke chips worden de AI-berekeningen uitgevoerd. Klassiek gebeurde dat op *central processing units* (CPUs), een industrie waarbinnen het Amerikaanse bedrijf Intel al lang de gigant is. Met de opkomst van de smartphones waren chips nodig die efficiënter energie gebruikten en het Amerikaanse bedrijf Qualcomm, dat ontwerpen van het Britse ARM gebruikt, werd daar leidend in. Zoals we in het eerste hoofdstuk zagen, bleek gaandeweg dat veel complexe berekeningen van AI beter uitgevoerd konden worden met zogenoemde *graphic processing units* (GPUs) die vooral in de gamingindustrie gebruikt werden en door bedrijven als Nvidia ontwikkeld worden.[503] Daarnaast bestaan inmiddels ook chips die speciaal ontworpen zijn voor een bepaald type berekeningen zoals machine learning-algoritmes. Deze technologie is zo specifiek en voor de sector van een dergelijk strategisch belang

502 Meer ondersteunende technologieën kunnen onderscheiden worden rondom productie en energietoevoer bijvoorbeeld. In een working paper in opdracht van de WRR onderzocht TNO het technisch ecosysteem uitgebreid (TNO 2021). De auteurs onderscheiden daarbinnen kerntechnologieën, complementaire technologieën en ondersteunende infrastructuur; zie hiervoor hoofdstuk 3 van het working paper.

503 Lee 2018: 96.

dat grote technologieplatformen die zelf ontwikkelen, zoals het geval is bij Googles TPUs en Microsofts FPGAs.[504]

Het is belangrijk om op te merken dat de ontwikkeling van deze ondersteunende technologie goeddeels gedomineerd wordt door bedrijven uit Silicon Valley. Nederland heeft door goede banden met de VS geen last van gebrek aan toegang hiertoe. In de handelsoorlog met China heeft de VS dat land de toegang tot kritische chiptechnologie ontzegd, wat ook effect heeft gehad op ons land. NXP, de vroegere halfgeleiderstak van Philips, kreeg in 2016 een overnamebod van het Amerikaanse Qualcomm, wat uiteindelijk niet doorging doordat China tegenwerkte. Ook het Nederlandse ASML speelt een belangrijke rol in de wereldwijde chipindustrie. Mede vanwege de essentiële rol voor AI is de chipindustrie voor Nederland dus van strategisch belang.

Een andere bron van rekenkracht zijn supercomputers. Die zijn voor veel alledaagse AI-toepassingen niet nodig, maar kunnen dat voor zeer complexe toepassingen in de toekomst wel worden. De Nationale Supercomputer Cartesius, ontwikkeld voor SURF, stond in 2018 op plek 218 van de 500 snelste supercomputers in de wereld. Cartesius wordt in 2021 vervangen door een nieuwe supercomputer van Lenovo, Snellius genaamd, die bijna tien keer zoveel rekenkracht heeft als zijn voorganger.[505]

Naast hardware is de andere grote ondersteunende technologie voor AI de grondstof waar ze op draait: data. Zeker de benaderingen van AI die nu dominant zijn, zoals deep learning, vereisen grote hoeveelheden data, veel meer dan de klassieke regelgebaseerde systemen. Belangrijk is dus ten eerste dat er data beschikbaar zijn. Hoe goed de algoritmes ook zijn, zonder relevante data kan AI niet werken. Bij de toepassing van AI gaat het er dus om te kijken of er data beschikbaar is en waar die zich bevindt. Het is niet voor niets dat AI als eerste veel is toegepast op consumentenplatformen die toegang hebben tot grote bronnen van data uit sociale media, zoekmachines en onlinewinkelgedrag.

Per sector verschilt het hoeveel data verzameld wordt en dat verschilt ook nog eens van land tot land. Neem de zorgsector. Een van de knelpunten voor AI in die sector is dat voor veel zaken de data beperkt aanwezig of bruikbaar is. Ziekenhuizen en instellingen gebruiken eigen systemen die niet met elkaar verenigbaar zijn en bepaalde data bestaat alleen handgeschreven of in papieren archieven. Die diversiteit hangt samen met het decentrale karakter van de Nederlandse zorg. In Frankrijk is die sector bijvoorbeeld anders georganiseerd,

504 Ding 2019: 23.
505 SURF, 8 februari 2021.

met universele systemen en gecentraliseerde databases. Dat is ook een van de redenen waarom die sector een van de pijlers is van de Franse AI-strategie.[506] De Raad voor Volksgezondheid en Samenleving (RVS) benadrukt in een verkenning over AI in de zorg ook het belang van goede data. Op de korte termijn dient er volgens de RVS aandacht te zijn voor de continuïteit van patiëntengegevens.[507]

Het gaat dus allereerst om de aanwezigheid van voldoende data, die bovendien voldoende kwalitatief en commensurabel moeten zijn, en voldoende toegankelijk. Door bedrijfsgeheimen, wetgeving, beroepsgeheim of simpelweg slechte systemen kan het zo zijn dat de grondstof om algoritmes op te trainen, data dus, beperkt aanwezig is. Ook deze situatie is herkenbaar in de Nederlandse zorg. AI-wetenschappers die verbonden zijn aan Nederlandse academische ziekenhuizen, zijn voor het trainen van hun algoritmes vaak aangewezen op Amerikaanse medische data. Dat komt omdat de Nederlandse data niet toegankelijk zijn of omdat het een moeilijk en onduidelijk proces is om die te verkrijgen, terwijl dat in de VS veel gemakkelijker is.

Dat brengt ons bij een ander punt over de noodzakelijke data: die moeten representatief zijn. Algoritmes getraind op de data van een bepaalde locatie geven geen garantie voor goede resultaten op andere locaties. Dat is duidelijk in het voorbeeld van de zorgdata. Populaties van verschillende landen kunnen andere genetische eigenschappen en levensstijlen hebben, waardoor conclusies niet één op één overdraagbaar zijn. Dat werd ook duidelijk met de uitbraak van de coronacrisis. Alhoewel er enerzijds mondiaal data werden gegenereerd, bleef het desondanks noodzakelijk om lokaal data te verzamelen, omdat er per land verschillen konden zijn in de ontwikkeling van het virus. Datzelfde geldt voor het domein van de mobiliteit. Verkeersborden, verkeersregels en stedelijke planning verschillen aanzienlijk van land tot land. Een autonoom voertuig dat in het ene land getraind is, werkt niet vanzelfsprekend ook goed in een ander land. Terwijl veel analyses op een mondiaal niveau plaatsvinden, is het dus ook vaak van belang om lokale data te gebruiken.

Ten dele liggen de uitdagingen voor de beschikbaarheid van goede data buiten het technische domein. Ze hebben te maken met de organisatie van een sector, wetgeving en normen en het opzetten van nieuwe systemen voor een goede datahuishouding. Voor een deel van die uitdagingen bestaan technische oplossingen. Een eerste voorbeeld zijn de reeds genoemde GANs, die op kunstmatige wijze nieuwe data genereren als er onvoldoende data beschikbaar is. Als de camera's van TomTom een straat filmen terwijl het regent, kunnen GANs data

506 France 2018.
507 RVS 2019.

genereren voor die straat zonder regen. Ook de techniek van *federated learning*, waarbij het algoritme naar de data wordt gestuurd in plaats van andersom, kan uitkomst bieden. Op deze manier kunnen algoritmes getraind worden met gevoelige data, zonder dat die gevoelige data daarvoor uit handen gegeven hoeven te worden of onrechtmatig worden gebruikt.

Tekstbox 5.2 – Ondersteunende technologie voor de zelfrijdende auto

Het bovengenoemde kader van ondersteunende technologieën kunnen we toepassen op de casus van de zelfrijdende auto. De aandacht gaat daarbij vaak uit naar de auto zelf en specifiek naar de intelligentie van het systeem dat die auto aandrijft. Een goedwerkende zelfrijdende auto vergt technisch echter veel meer dan alleen AI-software. Zowel op het gebied van hardware als op het gebied van data zijn er allerlei andere technologieën nodig om ervoor te zorgen dat AI hier kan werken. Er zijn vooralsnog verschillende technische benaderingen voor zelfrijdende auto's. In veel daarvan zijn de onderstaande technologieën essentieel.

Sensoren

Om *realtime* relevante data over het verkeer te verzamelen zijn goede sensoren nodig. Die hardware is nodig om de omgeving van de auto te kunnen scannen. Doordat veel camera's of andere scanners maar beperkt vooruit kunnen kijken, werken veel autonome systemen, gezien de noodzakelijke reactietijd, alleen op lage snelheden. Sensoren moeten in verschillende weersomstandigheden en omgevingen werken. Een fataal ongeluk met een Tesla vond bijvoorbeeld plaats omdat de auto een witte vrachtwagen niet goed kon onderscheiden van de witte horizon en daardoor niet remde. Gezien het potentiële gevaar worden enkele cruciale elementen in autonome voertuigen in drievoud geïnstalleerd.

Digitale kaarten

Andere belangrijke data voor zelfrijdende auto's komt van digitale kaarten die accuraat en up-to-date moeten zijn. Belangrijk voor dat laatste zijn bijvoorbeeld tijdelijke wegwerkzaamheden die de verkeerssituatie veranderen. Een ontwikkeling binnen digitale kaarten zijn de zogenoemde HD-kaarten (*high definition*) die door bedrijven als Tomtom en Waze worden ontwikkeld. Deze kaarten kunnen tot op de centimeter nauwkeurig de omgeving beschrijven, wat een grote vooruitgang in de dataverzameling betekent.

Rekenkracht
Zelfrijdende auto's moeten snel complexe berekeningen maken, wat rekenkracht van groot belang maakt. De Wet van Moore over de ontwikkeling van chips is in die zin ook belangrijk voor dit veld. Rond ongeveer het jaar 2005 werd het pas mogelijk om een 3D-kaart van een stad op de lokale geheugenschijf van een auto op te slaan.

I2C
Een volgende ondersteunende technologie betreft de fysieke en digitale infrastructuur. Hierbij gaat het om Infrastructure-to-Car (I2C) communicatie. Alhoewel verkeersborden en verkeerslichten voor mensen meestal heldere objecten zijn, is dat voor computers minder het geval. Een remedie hiervoor zou zijn om dat soort infrastructuur direct digitaal met de auto te laten communiceren zodat deze geen beelden hoeft te interpreteren ('Is dat een verkeerslicht of een remlicht van een vrachtwagen?'). Hiervoor is het wel nodig om chips in te bouwen in de fysieke infrastructuur. Nederland experimenteert hier al mee: minister Van Nieuwenhuizen van IenW kondigde in 2018 de aanleg van slimme verkeerslichten aan, Enschede heeft de SMART-app en er zijn experimenten in de regio Schiphol.

C2C
Een andere ondersteunende technologie is Car-to-Car-communicatie. Deze lost het probleem op dat het moeilijk is om de intenties van andere bestuurders in te schatten. Door geautomatiseerde communicatie tussen voertuigen zouden kop-staartbotsingen voorkomen kunnen worden, maar zouden auto's ook veel dichter op elkaar kunnen rijden wanneer ze simultaan kunnen remmen. Het in de vorige box genoemde 'platooning' is hier een voorbeeld van. Nederland experimenteert hiermee met de zogeheten 'Tulip corridors'.

Netwerk
Sommige benaderingen van zelfrijdende auto's plaatsen de rekenkracht in de auto zelf terwijl andere meer intelligentie over het netwerk plaatsen. Wanneer het gaat om communicatie buiten de auto (I2C of C2C), is een goed digitaal netwerk een vereiste. Specifiek de toekomstige 5G-netwerken kunnen hierbij een belangrijke rol spelen. Daar gaan we bij de volgende box over emergente technologieën verder op in.

Techniek in een enveloppe

Een relevant concept rondom het vraagstuk van ondersteunende technologieën is dat van *enveloping*: het beter laten functioneren van een nieuwe technologie, niet door verbeteringen aan die technologie zelf maar door aanpassingen aan de omgeving waarin de technologie opereert. De noodzaak daarvan bij AI wordt duidelijk aan de hand van een aantal voorbeelden waar dat niet gebeurd is, met grote problemen tot gevolg. In haar boek *Artificial Unintelligence* noemt Meredith Broussard een voorbeeld uit het onderwijs. Als oplossing voor de slechte beschikbaarheid van lesboeken in veel Amerikaanse wijken werd de introductie van *e-education* als een oplossing gezien, waarbij kinderen door een enkele investering hun lesmateriaal via telefoons of tablets konden krijgen. Hierbij werd echter het hele ecosysteem eromheen vergeten dat nodig is om deze toepassing te laten functioneren. Denk bijvoorbeeld aan de infrastructuurkosten. De computers vergen onderhoud en allerlei andere diensten, zoals telefoonlijnen en email om kinderen bij problemen te kunnen helpen. Zeker in oude scholen moet gekeken worden of de elektra wel geschikt is om al die computers aan te sluiten en op te laden. Het wifinetwerk moet altijd en overal op school goed werken. Toegangscodes moeten worden opgezet, evenals veilige leeromgevingen die de privacyregels niet schenden. Voor alle gebruikers moeten identificatiebewijzen worden aangemaakt, oud-leerlingen moeten verwijderd worden, licenties voor digitale boeken moeten geregeld worden. De digitalisering van het onderwijs vindt dus niet plaats door simpelweg computers uit te delen.

Een ander voorbeeld van onvoldoende aansluiting tussen de omgeving en een AI-systeem betreft zelfrijdende auto's. Broussard merkt op dat, ondanks dat daarvoor online grote trainingsets bestaan bij arXiv en GitHub, die data onvoldoende vreemde situaties bevatten en dat het eigenlijk ook onmogelijk is om data te hebben voor alles wat op de weg kan gebeuren. Kinderen die een nieuw spel spelen en daarbij onvoorspelbare bewegingen langs de weg uitvoeren of een persoon die rondjes rent achter een dier aan, zijn situaties die mensen snel kunnen duiden maar autonome voertuigen in de war brengen. In Australië worden autonome vrachtwagens ingezet in de mijnbouw. Dit zijn relatief gecontroleerde omgevingen waar mensenwerk al in hoge mate geautomatiseerd is, waardoor deze voertuigen zonder grote gevaren kunnen rijden.[508] In tekstbox 5.3 gaan we uitgebreider in op *enveloping* van de zelfrijdende auto.

508 Agrawal et al. 2018: 113.

Tekstbox 5.3 – De zelfrijdende auto en de fysieke omgeving

De Nederlandse infrastructuur
Voor de zelfrijdende auto maakt KPMG jaarlijks de Autonomous Vehicle Readiness Index. Nederland voert daarin al twee jaar de lijst aan op het gebied van infrastructuur. Dat komt door de hoge kwaliteit van het wegdek, het wegontwerp en signalisatie door middel van borden (KPMG, Index 2019).

Het wegdek moet niet alleen van goede kwaliteit zijn, maar ook goed onderhouden worden. De belijning moet goed zichtbaar blijven, ook onder slechte weersomstandigheden en na slijtage. Verder is het belangrijk dat het wegdek eenduidig is. Een uitdaging daarbij zijn (spoed)werkzaamheden. Daarvoor wordt de belijning vaak zwart gemaakt of weggewist en vervangen door een tijdelijke gele belijning. Dat kan voor zelfrijdende auto's een probleem zijn omdat ze dan twee conflicterende bronnen van data binnenkrijgen. Een uitdaging bij het wegontwerp zijn specifiek Nederlandse constructies als tapercon-structies, weefvakken en spitsstroken. De verschillende doorgetrokken lijnen kunnen ook verwarrend zijn voor de systemen van autonome voertuigen. Rijkswaterstaat en RDW voeren gesprekken over mogelijke aanpassingen aan de weginrichting in dit kader.

In plaats van AI-toepassingen net zo lang te verbeteren tot ze op het gewenste niveau kunnen presteren, gaat *enveloping* over het aanpassen van de omgeving. Door die omgeving beter 'leesbaar' te maken voor AI, kan het systeem gemakkelijker met die omgeving omgaan zonder alles te kunnen wat een mens ook kan. Vergelijk dit met de Wright Brothers die het eerste vliegtuig ontwikkelden (Broussard 2019: 131). Voor die tijd dachten mensen dat vliegen klappende vleugels vereist, zoals bij vogels; een proces dat wetenschappers pas sinds kort kunnen imiteren. Het vliegtuig van de Wrights imiteerde geen vogel, maar deed iets anders. Zo ook hoeft een zelfrijdende auto ook niet alles te kunnen wat een menselijke bestuurder kan als de omgeving aangepast is.

De uitrol van zelfrijdende auto's
Wat betekent dit nu voor de toepassing van zelfrijdende auto's? Deze zullen waarschijnlijk, net als de eerdergenoemde robots, eerst toegepast worden in eenduidige en voorspelbare omgevingen. Te denken valt aan bussen op bedrijventerreinen, op de landingsbanen van vliegvelden of op relatief afgesloten gebieden met weinig verkeer zoals golfbanen en verzorgingstehuizen.

Mogelijk kunnen op termijn zelfrijdende auto's op de snelweg goed functioneren, maar de binnenstad is veel uitdagender. Dat zou kunnen betekenen dat op specifieke locaties mensen of goederen van transportmiddel veranderen en na de snelweg overstappen van een zelfrijdende auto op een vervoermiddel met chauffeur. Dat vergt een geïntegreerd transportsysteem, zoals het huidige systeem waarin treinen en bussen met elkaar verbonden worden.

Het is niet uit te sluiten dat rigoureuzere maatregelen in de infrastructuur nodig zijn om zelfrijdende auto's echt verantwoord de weg op te kunnen laten gaan. Neem het volgende vraagstuk. In moeilijke omgevingen moet een chauffeur het rijden van een AI-systeem over kunnen nemen. Onderzoek wijst uit dat het ongeveer 20 seconden kost om daartoe in staat te zijn. Het is dus plausibeler om te kijken in welke situaties een AI-systeem de auto kan besturen en wanneer de mens dat moet doen. Daarvoor is in Nederland gekeken naar een Operational Design Domain (ODD): het gebied waar een zelfrijdende auto goed zou kunnen functioneren. Slechts een aantal stroken van een paar kilometer en een stuk over de Afsluitdijk voldeden aan alle eisen van het ODD.

Rigoureuzere maatregelen zouden de bouw van aparte banen of wegen kunnen betekenen, waar geen andere voertuigen mogen komen. Dat gebeurde ook met de oorspronkelijke auto. Het Rijkswegenplan van 1927 stelde hoofdwegen in, bedoeld voor alleen gemotoriseerd verkeer. Fietsers werden naar een ander wegennet verplaatst. Ook de zelfrijdende auto zou kunnen beginnen in een hoog gecontroleerd gebied en van daaruit stapsgewijs met aanpassingen in andere gebieden kunnen gaan werken.

Enveloping verwijst dus naar de aanpassing van de fysieke omgeving om AI-toepassingen goed te laten functioneren. Wat betekenen die noodzakelijke aanpassingen voor onze verwachtingen van AI-toepassingen? Ten eerste dat, terwijl de aandacht vaak uitgaat naar de vaardigheden van autonome voertuigen, robots of drones, ook veel vooruitgang bij die toepassingen geboekt kan worden door de omgeving aan te passen. Om drones ooit echt goederen te kunnen laten bezorgen is standaardisering van bijvoorbeeld landingslocaties nodig of een nieuw soort brievenbussen, veilige routes waardoor deze voertuigen mensen niet kunnen raken en systemen om de identiteit van een geadresseerde te verifiëren.

Ten tweede betekent het idee van *enveloping* dat door de vereisten aan de omgeving veel AI-toepassingen eerst en vooral ingezet kunnen worden in specifieke omgevingen en daarna pas uitgebreid kunnen worden. Neem het

voorbeeld van robots. Alhoewel de fantasie van een robotassistent in huis al heel lang bestaat, is dat waarschijnlijk een van de laatste plaatsen waar die terecht zal komen. De thuisomgeving heeft namelijk een zeer hoge complexiteit van mogelijke taken (opruimen, een dinerfeestje organiseren, huiswerkbegeleiding, persoonlijke hygiëne) en heel veel onvoorspelbare variabelen (kleine kinderen, huisdieren, fragiele objecten). Zie tekstbox 5.4.

Tekstbox 5.4 – Stofzuigers en huizen

Roomba is een autonome stofzuiger in de vorm van een platte ronde schijf gemaakt door het bedrijf iRobot, die het huis schoonmaakt op basis van camera's en eerdere ervaring. Een probleem met de omgeving is dat het apparaat door de ronde vorm de hoeken van het huis niet goed kan bereiken. Een groter probleem dient zich echter aan bij de eigenaren van huisdieren. Voor stof en vuil werkt de Roomba goed, maar niet wanneer het ontlasting betreft. Die verspreidt de stofzuiger namelijk door het hele huis, een fenomeen dat inmiddels de *poopocalypse* wordt genoemd.

Over het probleem van de hoekige vormen heeft Luciano Floridi een interessante bespiegeling die relevant is voor het idee van enveloping. Hij stelt als gedachtenexperiment voor dat we om de Roomba beter zijn werk te kunnen laten doen, in de toekomst in ronde kamers zouden kunnen gaan wonen. Een dergelijke aanpassing van de omgeving vergroot de effectiviteit van de stofzuiger aanzienlijk. Veel mensen zullen geneigd zijn om kritisch te zijn over het aanpassen van ons leven aan de technologie in plaats van andersom. Floridi vraagt vervolgens echter of we dat nu niet al doen: wonen we niet onder andere in vierkante kamers omdat dat deze beste passen bij rechthoekige bakstenen?[509]

De toepassing van robots zal volgens Stuart Russell daarom in stappen via andere domeinen verlopen. Initieel bijvoorbeeld in warenhuizen, zoals nu al gebeurt met de robots van Amazon. De taken zijn simpel en eenduidig – 'breng X naar Y' – en de omgevingen zijn gecontroleerd, waardoor de robots efficiënt hun werk kunnen doen.[510] Een volgende stap is de toepassing in andere commerciële omgevingen, zoals landbouw en constructie, waar de taken en objecten redelijk voorspelbaar zijn. Daarna komt het vakkenvullen en kleding

509 Floridi 2014: 150-1.

510 Voor een demonstratie van hoe dat werkt, zie het filmpje van Kiva Robots: Amazon (YouTube, 24 juli 2017).

sorteren in de detailhandel. Wanneer robots het huis in gaan, zal dat eerst zijn voor ouderen en mensen met een beperking waarvoor de robot specifieke handelingen kan verrichten. Ver daarna pas is zoiets als een universele robotbutler denkbaar.[511]

Die fasering is vooral van belang als het gebruik van AI levens op het spel kan zetten, zoals met robots in huis of voertuigen in de binnenstad. Toepassingen waarbij minder op het spel staat, kunnen al eerder in minder gecontroleerde omgevingen terechtkomen, nog voordat ze goed functioneren. Denk hierbij aan de virtuele assistenten als Alexa en Siri in de thuisomgeving. Het is duidelijk dat we daar nog niet normaal mee kunnen converseren. Ze vergen van ons dat we onze woorden op een specifieke manier uitspreken en in een bepaalde zinsopbouw, zodat het programma ons kan begrijpen. En dan is nog steeds niet gegarandeerd dat we het juiste antwoord krijgen. Omdat ze voor beperkte zaken net goed genoeg zijn ("Waar is de dichtstbijzijnde fietsenmaker?") en als gadgets mensen aanspreken, vinden we ze nu toch al in veel huishoudens. Daar kunnen ze ook meteen veel data verzamelen waarmee ze later nuttiger kunnen worden.

Dit brengt ons bij een laatste implicatie. AI-systemen kunnen sneller functioneren in nieuwgebouwde omgevingen die met het oog op deze toepassingen zijn gemaakt, dan wanneer zij in bestaande omgevingen ingepast worden. Dat kan een reden zijn om nieuwe infrastructuur aan te leggen of om bij de aanleg van nieuwe infrastructuur met de AI-toepassing rekening te houden. Daarom kan een land als China grote voortgang boeken met AI-toepassingen. Door snelle urbanisering bouwt het land in rap tempo nieuwe steden en wijken, die van meet af aan ingericht kunnen worden voor bijvoorbeeld autonome voertuigen.

Kernpunten – Het technische ecosysteem: ondersteunende technologie

- Onderdeel van het technische ecosysteem zijn ondersteunende technologieën.
- AI vergt ondersteunende hardware in de vorm van netwerken, chiptechnologie en supercomputers.
- AI vergt daarnaast als grondstof data die breed, kwalitatief hoog, commensurabel, beschikbaar en representatief moet zijn.
- Om een nieuwe technologie succesvol te implementeren is enveloping een effectieve maar onderschatte strategie: het aanpassen van de omgeving aan een technologie zodat die beter kan functioneren.

511 Russell 2019: 74.

Emergente technologieën

Ondersteunende technologieën laten zien dat een nieuwe technologie onderdeel is van een breder technisch ecosysteem. Dat brengt complexiteit en onzekerheid voor het gebruik van een nieuwe technologie met zich mee. Dit geldt des te meer voor emergente technologieën. Terwijl ondersteunende technologieën namelijk vanaf het begin al aanwezig zijn bij de nieuwe technologie, gaat het hier om technologieën die er aanvankelijk los van staan. Emergente technologieën worden parallel of op een later moment elders ontwikkeld, waarna ze aan de betreffende technologie verbonden worden. Dat proces van inbedding in emergente technologieën is nog moeilijker te voorzien dan de inbedding in ondersteunende technologieën. Elektriciteit had in huis aanvankelijk maar beperkte toepassing, namelijk als bron van licht. Later werd het verbonden met andere innovaties in het domein van de elektronica. Niemand had kunnen voorzien op welke manier dat tot de elektrificatie van het huishouden zou leiden in de vorm van allerlei huishoudelijke apparatuur.

Die onzekerheid over emergente technologieën geldt ook voor AI. De toepassing ervan in de samenleving is pas van korte duur. Het valt niet te voorzien met welke andere technologieën AI verbonden zou kunnen worden die nu parallel ontstaan, laat staan technologieën die nog in de kinderschoenen staan. Desondanks kan een dergelijke verbinding wel een enorme impuls aan AI geven of de toepassing ervan een bepaalde kant op duwen. Het is dus belangrijk om hier oog voor te hebben. Daarom bespreken we hier kort een aantal emergente technologieën die met AI verbonden kunnen raken. We beginnen met de meest volwassen parallelle technologieën en werken toe naar de prillere.

Eerder al bespraken we netwerktechnologie als ondersteunende technologie. Die technologie staat niet stil en een volgende generatie die opkomt, is 5G. Die ontwikkeling zorgt niet alleen voor een sprong in snelheid, maar kent ook een andere infrastructuur en maakt andere toepassingen mogelijk. Momenteel worden experimenten met de uitrol van 5G gedaan en die zullen dan ook effect hebben op de mogelijkheden van AI (zie tekstbox 5.5).

Een andere technologie die zich parallel ontwikkelt en al in een ver stadium is, is het zogenoemde Internet of Things (IoT). Hierbij gaat het over het plaatsen van sensoren en chips in allerlei objecten in de fysieke omgeving waarmee ze met het internet verbonden worden. Ontwikkelingen in nanotechnologie die steeds kleinere en goedkopere hardware mogelijk maken zijn een belangrijke drijfveer van deze ontwikkeling. Het IoT gaat over het verbinden van wegen en verkeerslichten, maar ook dijken, broodroosters, speelgoed, speakers, fabrieken, koelkasten, kleding en zelfs dieren en onze lichamen.

Volgens het Amerikaanse bedrijf Cisco dat veel van die hardware maakt, bereikten we in 2008-2009 het keerpunt waarop meer objecten met het internet verbonden waren dan mensen. De International Data Corporation schat in dat er in 2025 wereldwijd meer dan veertig miljard IoT-apparaten zijn. De technologie van IoT zal in toenemende mate met AI verbonden worden. Een van de bouwstenen van AI zijn immers data. AI kreeg de afgelopen jaren een sterke impuls van de immense hoeveelheid data die mensen achterlaten op het internet en het IoT zal daar nieuwe data over de fysieke wereld aan toevoegen. Het is daarmee een belangrijke factor voor nieuwe AI-toepassingen.

Tekstbox 5.5 – De zelfrijdende auto en emergente technologieën

In de vorige box beschreven we hoe de zelfrijdende auto een goed digitaal netwerk vereist. De volgende generatie daarvan, 5G, zou daar een belangrijke rol bij kunnen spelen. Meer nog dan in andere gebieden is snelheid hier namelijk van belang. Te laat remmen of een hapering in de verbinding kan het verschil betekenen tussen leven en dood. De veel hogere snelheid van 5G is daarom belangrijk, maar ook de veel lagere latentie van deze netwerken (de duur tussen het verzenden en ontvangen van een signaal). De transitie naar 3G en 4G maakte het mogelijk om op smartphones video's en films te streamen. Het eerdere netwerk was hiervoor simpelweg te zwak. Op eenzelfde manier kan volgens experts 5G een voorwaarde zijn voor goedwerkende autonome voertuigen.

Een andere emergente technologie voor de zelfrijdende auto is elektrisch rijden. Het is niet voor niets dat veel elektrische auto's geavanceerde vormen van automatisering kennen (Tesla, Nissan Leaf, Volvo). Beide technologieën vereisen namelijk een uitgebreide automatische transmissie. De bijbehorende nieuwe infrastructuur voor elektrische auto's, zoals oplaadpunten, kan daarmee ook verbonden worden aan die voor autonome voertuigen.

Een volgende emergente technologie zijn cryptocurrencies en de onderliggende technologie van blockchain. Rondom die technologieën is de laatste jaren veel te doen geweest, zeker rondom de bekendste cryptocurrency Bitcoin. Het koersverloop daarvan toont de hypegevoeligheid van deze technologie. Op dit moment valt nog niet te zeggen op welke schaal en op wat voor manier deze technologieën toegepast zullen worden. Tegelijkertijd bieden zij wel enorme mogelijkheden, juist in combinatie met andere technologieën. Cryptocurrencies maken, voortbordurend op de blockchaintechnologie, bijvoorbeeld een decentraal systeem van betalingen mogelijk. Gekoppeld aan AI valt te denken aan de detectie van het gebruik van iemands intellectueel

eigendom, een lied of artikel, waarna de maker automatisch betaald wordt.[512] Ook is goed voorstelbaar dat fietsen en huizen, verbonden aan het IoT, op platformen als Airbnb digitaal geopend en vergrendeld worden voor de duur dat iemand ervoor betaald heeft. Toegang tot bepaalde objecten of locaties zou via de digitale weg in de fysieke werkelijkheid gegrift kunnen worden.

De achterliggende technologie van blockchain kan niet alleen voor betalingen gebruikt worden, maar ook om allerlei andere transacties op decentrale wijze vorm te geven. Een mogelijk voordeel ervan is de grotere veiligheid en lagere afhankelijkheid van centrale spelers of databases. Hoewel dit laatste ook zijn nadelen kent, kunnen deze eigenschappen de barrières voor allerlei AI-toepassingen verlagen. AI en blockchain ontmoeten elkaar in zogenoemde 'DAO's' (*distributed autonomous organizations*), organisaties die niet uit mensen, maar uit regels en contracten bestaan die automatisch beslissingen kunnen nemen.

Een nog minder volwassen technologie is *quantum computing*. Dat is een technologie die belooft de rekenkracht van computers immens te vergroten. Simpel gezegd werken traditionele computers met bits in een binaire logica van enen en nullen. Quantumcomputers werken met *quantum bits* of *qubits*, die meerdere staten tegelijkertijd kunnen aannemen, waardoor het aantal mogelijke berekeningen veel hoger ligt.[513] In plaats van met brute rekenkracht berekeningen te maken, kunnen deze computers alle mogelijke configuraties meteen representeren.

De technologie is nog in ontwikkeling, er zijn verschillende benaderingen en de technologie presteert nog niet beter dan reguliere computers. Wanneer dat echter wel gebeurt, het moment van *'quantum supremacy'*, wordt volgens experts een immense sprong gemaakt. Alle encryptie die op grote rekenkracht berust, zou dan teniet worden gedaan alsof iemand in een keer de sleutel voor alle kluizen in de wereld in handen krijgt. Daarom investeren landen als de VS en China flink in deze technologie. Ook Europa is actief op dit gebied, en Nederland is met een onderzoeksgroep in Delft een wereldspeler in *quantum computing*.[514]

Ondanks dat *quantum computing* nog in de kinderschoenen staat, is het goed voor te stellen dat de technologie een revolutie zou zijn voor het gebruik van AI. De groei in rekenkracht is, zoals we hebben gezien, een van de pijlers daarvan. De combinatie van *quantum computing* en AI zou zeer complexe vraagstukken van data-analyse of wetenschappelijk onderzoek naar medicijnen bijvoorbeeld

512 Greenfield 2017: 133.
513 Vermaas et al. 2019.
514 Voor een overzicht van de status van het onderzoek in Delft in 2019, zie: QuTech 2019.

een grote impuls kunnen geven. Niet voor niets doen partijen als Google al onderzoek naar '*quantum AI*'.

Een systeemtechnologie als AI werkt altijd in een technisch ecosysteem, zo blijkt uit de beschrijving van ondersteunende en emergente technologieën. Dat brengt veel complexiteit en onvoorspelbaarheid met zich mee. Niet alleen ontwikkelingen in de technologie zelf, maar ook elders in het ecosysteem kunnen faciliterend of verhinderend werken voor de toepassing ervan. Dat verklaart waarom sommige toepassingen al heel voorstelbaar zijn en in een laboratoriumsituatie goed werken – zoals zelfrijdende auto's op een racebaan –, maar nog ver verkeren van implementatie in het normale leven. Anderzijds kunnen verbeteringen elders plots tot een impuls leiden waardoor wat een stagnerende technologie leek, opeens grote vooruitgang kan boeken.

Hoe staat het in Nederland met een benadering voor AI die rekening houdt met ondersteunende en emergente technologieën? Nadat het Strategisch Actieplan voor AI was gepresenteerd, zijn de actiepunten vervolgens opgenomen in de bredere Digitaliseringsstrategie van de overheid. Dat is belangrijk, want daarmee wordt de verwevenheid met andere technologieën duidelijk. Technologieën als AI moeten niet in aparte silo's behandeld worden. Een volgende stap is om in die strategie meer aandacht te schenken aan opkomende technologieën als het IoT, 5G, blockchain en *quantum computing*. De grote les van emergente technologieën is dat een technologie een enorme impuls kan krijgen door innovatie in een hele andere technologie. Voor de toekomst van AI is het dus ook belangrijk om scherper zicht te hebben op nieuwe ontwikkelingen in andere technologieën en ook daarin te investeren.

Kernpunten – Het technische ecosysteem: emergente technologie

- Emergente technologieën zijn technologieën die aanvankelijk los van elkaar staan, maar onderling wel een groot effect kunnen hebben op verdere ontwikkeling als ze aan elkaar gekoppeld worden.
- Voor AI lijken 5G, IoT, blockchain en *quantum computing* kandidaten te zijn als emergente technologieën.
- Het verloop van die andere technologieën is niet te voorspellen, maar het is prudent om ze mee te nemen in de ambities en strategieën rondom AI.
- De beide dimensies van het technische ecosysteem met ondersteunende en emergente technologieën verklaren waarom een technologie die ogenschijnlijk gereed is voor gebruik, pas veel later in de praktijk tot wasdom komt, maar ook waarom praktische toepassing plotseling in een versnelling kan komen.

5.2 Het sociale ecosysteem

De macro-economische context

De inbedding van AI in het sociale ecosysteem brengt op macroniveau, dat van de economie, twee grote vraagstukken met zich mee. Het eerste gaat over het effect op de werkgelegenheid en specifiek 'technologische werkloosheid', het tweede gaat over het fenomeen van de productiviteitsparadox. Beide vraagstukken betreffen langetermijneffecten van AI waarvoor geldt dat het op dit moment onmogelijk is er zicht op te hebben. Tegelijkertijd leert de geschiedenis van systeemtechnologieën ons wel om op een bepaalde manier naar die vraagstukken te kijken en biedt ze ons handvatten om de bijbehorende fenomenen in goede banen te leiden.

Het eerste vraagstuk dat herhaaldelijk terugkomt in de geschiedenis van technologische revoluties, gaat over de vrees dat banen massaal zullen verdwijnen, waardoor grote groepen mensen niet meer in hun levensonderhoud kunnen voorzien. Er zijn ook positieve varianten van dit idee omdat de technologie ons bevrijdt van saai, gevaarlijk en hard werk en ruimte creëert voor andere meer betekenisvolle activiteiten. Dit idee komen we al tegen bij Karl Marx. In de communistische eindstaat zou de mens zich bezig kunnen houden met jagen, vissen en kritieken schrijven.[515] En de econoom John Maynard Keynes beschreef in 1930 een toekomst waarin we maar een paar uur per dag zouden hoeven te werken.[516]

Terugkijkend wordt er vaak gekscherend gedaan over de angst voor het verdwijnen van werk. Al eeuwen maken mensen zich zorgen over de effecten van ploegen, machines en pinautomaten, en desondanks zijn we niet massaal zonder werk komen te zitten. Symbolisch voor die overtrokken angst zijn de in hoofdstuk 3 besproken Luddieten. Deze wevers waren bang werkloos te worden door de Industriële Revolutie. In plaats daarvan leidde die revolutie tot allerlei nieuwe banen. Toch hadden de Luddieten wel degelijk een punt.[517] Voor hier is het belangrijk om op te merken dat de historisch constante aanwezigheid van werk niet zomaar op de toekomst geprojecteerd kan worden. Zeker omdat een aantal auteurs betoogt dat hedendaagse technologie als AI anders is dan de technologie van het verleden.

Een beroemd boek in dit genre is *The Second Machine Age* van Erik Brynjolfsson en Andrew McAfee, waarin zij een hedendaagse digitale technologie als AI tevens als een GPT beschouwen. Waar het eerste machinetijdperk – de

515 Marx en Engels 2010 [1932].
516 Keynes 1930.
517 Tielbeke, 16 mei 2018.

Industriële Revolutie – volgens hen complementair was aan menselijk werk, zien zij de technologieën in het tweede machinetijdperk als substitutief. Het eerste machinetijdperk verving spierkracht en leidde tot een proces van '*deskilling*' waarbij allerlei complexe vaardigheden van ambachtslieden werden opgebroken in simpele taken die veel arbeiders in fabrieken konden uitvoeren. Het huidige machinetijdperk vervangt echter ook onze denkkracht en zal volgens Brynjolfsson en McAfee in rap tempo menselijke arbeid overbodig maken. Belangrijk bewijs dat zij hiervoor presenteren, is wat zij de *'spread'* noemen: de toenemende ongelijkheid bij hedendaagse technologie doordat de lonen van grote groepen werknemers achterblijven.[518]

Echt grote zorgen over de toekomt van werk als gevolg van AI ontstonden echter na een studie van twee wetenschappers uit Oxford. Carl Benedikt Frey en Michael A. Osborne voorspelden in 2013 dat 47 procent van de Amerikaanse banen in de komende tien tot twintig jaar geautomatiseerd zou kunnen worden. In hun studie stelden zij dat het technisch mogelijk zou zijn om die banen te automatiseren, en niet dat dat in die periode ook daadwerkelijk zou gebeuren. Desalniettemin leidde dit paper wereldwijd onmiddellijk tot veel ophef.

In 2016 suggereerden economen van de OESO dat de situatie niet zo dramatisch was. Zij kwamen erop uit dat 9 procent van de banen onder druk staat. Deze uitkomst had ermee te maken dat ze de nadruk niet legden op banen maar op taken. Hoewel veel taken geautomatiseerd kunnen worden, geldt dat echter niet tegelijkertijd voor de banen waar ze onderdeel van zijn. In 2017 schatte PwC dat 38 procent van de Amerikaanse banen een groot risico liep om in de vroege jaren dertig geautomatiseerd te zijn. Volgens een McKinsey-studie is op dit moment al 50 procent van de wereldwijde banen te automatiseren.

Wat in deze context nog relevant is om te vermelden, is dat AI het effect van automatisering anders maakt dan in het verleden. Dat heeft te maken met de eerdergenoemde Moravec-paradox: sommige zaken die voor ons moeilijk zijn, zijn voor computers gemakkelijk en vice versa. Terwijl eerdere automatisering vooral fysieke fabrieksarbeid verving, raakt AI verschillende intellectuele en conceptuele vaardigheden van mensen en dus administratieve, financiële en andere 'witteboordenbanen.[519] Daarentegen zullen computers de motorische vaardigheden van kappers, chauffeurs of schoonmakers voorlopig nog niet kunnen vervangen.

518 Brynjolfsson en McAfee 2014.
519 Lee 2018: 166.

Als gevolg van deze scenario's worden allerlei oplossingen bedacht om met het verlies van werk om te gaan, ook vanuit Silicon Valley waar deze veranderingen hun oorsprong hebben. Larry Page van Google stelde bijvoorbeeld voor om korter te gaan werken, waardoor meer mensen het bestaande aantal banen kunnen delen. Een ander veel gehoord voorstel is het zogenoemde universele basisinkomen. Wat de uiteindelijke effecten van AI op de arbeidsmarkt zullen zijn, is nu niet te voorzien. In andere studies is de WRR uitgebreider op dit vraagstuk ingegaan.[520] In het verlengde daarvan plaatsen wij hier een aantal vraagtekens bij het idee dat de meeste banen zullen verdwijnen.

Ten eerste toont de geschiedenis van systeemtechnologieën dat deze angsten steeds terugkeren. Mensen zien eerder de banen die verdwijnen dan de nieuwe banen die ontstaan. Dat geldt ook voor AI nu. Ondanks de verwachtingen in voornoemde studies tonen de arbeidsmarktcijfers niet aan dat het aantal banen structureel afneemt, en een aantal sectoren kent zelfs grote tekorten.

Een ander belangrijk punt in dit verband is dat het nog maar de vraag is wat de oorzaak is van specifieke fenomenen zoals de ongelijkheid of 'spread' in lonen. Terwijl Brynjolfsson en McAfee die oorzaak evenals Kai-Fu Lee toeschrijven aan de aard van de technologie, is dat maar een deel van het verhaal. Technologieën als AI zullen er zeker aan bijdragen dat banen verdwijnen in het middensegment en kapitaal zich concentreert aan de bovenkant. Tegelijkertijd zijn er veel andere factoren die hier grote invloed op hebben. Een daarvan is neoliberaal beleid, waardoor de positie van georganiseerde arbeid is verzwakt, sociale vangnetten zijn verminderd en de belastingdruk minder nivellerend is geworden. Dat heeft ook bijgedragen aan de stagnatie van veel lonen. Globalisering is een andere factor. Door opkomende landen in vooral Azië is een grote hoeveelheid goedkope arbeid aan de wereldwijde arbeidsmarkt toegevoegd, met negatieve consequenties voor de lonen.

Bovendien, zoals de WRR in *Het Betere Werk* benadrukt en ook in het voorgaande hoofdstuk aan de orde kwam, is het effect van technologie allerminst gedetermineerd.[521] Achter de manier waarop technologie wordt ingezet en de effecten die die inzet heeft op de werkgelegenheid, zitten economische en politieke keuzes. Wij moeten dus terughoudend zijn met de bewering dat de effecten die wij nu waarnemen, grotendeels inherent zijn aan technologieën als AI. Sterker nog, het idee van de maatschappelijke inbedding van AI gaat juist over het bewuster omgaan met de inzet ervan en het borgen van de publieke belangen daarbij.

---

520 WRR 2013, 2015 en 2020.
521 WRR 2020.

Dat we vraagtekens kunnen plaatsen bij het idee dat de meeste banen zullen verdwijnen, betekent niet dat we geen aandacht moeten hebben voor de effecten van AI op de arbeidsmarkt. Veel banen zullen blijven bestaan, maar het is goed mogelijk dat er een groot verschil ontstaat tussen wat die banen straks vereisen als AI een grotere rol gaat spelen en de vaardigheden die mensen nu hebben. Dat is het echte vraagstuk van het effect van AI op de arbeidsmarkt. 'Technologisering' is een van de fundamentele veranderingen die nu in de wereld van werk plaatsvindt en vraagt aanpassingen van mensen.[522] Ook dat strookt met de lessen van systeemtechnologieën. Terwijl er na de Industriële Revolutie allerlei nieuwe banen ontstonden, was het overgangsproces moeizaam en pijnlijk. Het ging gepaard met werkloosheid, ongelukken en ellende in de overvolle binnensteden van Europa. De nieuwe werkcondities misten bovendien nog adequate regels en kaders. Denk aan kinderarbeid en de exploitatie van arbeiders in de negentiende eeuw, verbeeld in de boeken van Charles Dickens. Mensen moesten nieuwe vaardigheden leren en misstanden op het werk moesten geadresseerd worden.

Ook nu is er dus een tweeledige opgave om AI in te bedden in de wereld van werk. Ten eerste dient de discussie verlegd te worden van het verdwijnen van banen naar de transformatie van banen. Dat betekent niet meer uitgaan van het idee dat de mens het moet opnemen tegen de machine. Exemplarisch hiervoor is een opmerking van de Nederlandse schaakgrootmeester Jan Hein Donner op de vraag hoe hij zich op een wedstrijd tegen een computer zou voorbereiden: "Ik zou een hamer meenemen".[523]

In plaats van 'mens-tegen-machine' zou de aandacht uit moeten gaan naar de combinatie 'mens-met-machine'. AI gaat dan niet primair om het vervangen van menselijke intelligentie, maar om het vergroten daarvan, ook wel '*intelligence augmentation* (IA)' genoemd. Volgens Frank Pasquale is, in tegenstelling tot allerlei alarmistische verhalen ("*software is eating the world*"), het daadwerkelijke effect van AI dat de technologie mensen ondersteunt en versterkt in het doen van hun werk.[524] De beroemde AI-wetenschapper Geoffrey Hinton zei ooit dat we nu moeten stoppen met het opleiden van radiologen. De auteurs van *Prediction Machines* laten echter zien dat AI radiologen kan helpen in hun werk en dat er ten minste vijf rollen zijn die op dit moment niet door AI vervangen kunnen worden.[525]

522 WRR 2020.
523 Brynjolfsson en McAfee 2019: 189.
524 Pasquale 2020: 13-14.
525 Agrawal et al. 2018: 145-8.

Een goede combinatie tussen mens en machine vereist ervaring met en kennis van AI. Specifiek relevant is praktische kennis. In deze fase van maatschappelijke inbedding komt het er, net als eerder bij elektriciteit, op aan om te bedenken hoe allerlei domeinen, apparaten en praktijken met de nieuwe technologie verrijkt kunnen worden en hoe dat verantwoord kan. Wat dat concreet voor individuen in hun werk betekent, bespreken we hierna.

Ook het effect van AI op de arbeidsomstandigheden vraagt meer aandacht. Net als tijdens de Industriële Revolutie vinden nu allerlei misstanden plaats in de door nieuwe technologie mogelijk gemaakte banen. Illustratief zijn de werkomstandigheden en rechten van werknemers op platformen als Uber en Deliveroo of de arbeiders in de sorteercentra van Amazon, wiens toiletpauzes minutieus bijgehouden worden en arbeidsomstandigheden worden bepaald door de 'rate' die dynamisch doelstellingen formuleert. Tegelijkertijd verspreidt die trend waarbij AI wordt ingezet om werknemers meer te surveilleren, zich rap door de hele economie. AI Now documenteerde allerlei manieren waarop werknemers middels technologie onder erbarmelijke omstandigheden werken, van migrantenarbeid in de landbouw tot sensoren die werknemers vertellen hoe ze moeten lopen en wat ze moeten doen.[526] In Nederland heeft het Rathenau Instituut gewaarschuwd voor de groeiende trend van digitale monitoring op de werkvloer.[527] Hoewel AI mensen dus niet op de korte termijn massaal zal vervangen, zal het sommige banen deels automatiseren, transformeren en de werkomstandigheden beïnvloeden.

Het tweede grote macro-economische vraagstuk rondom systeemtechnologieën is de eerdergenoemde productiviteitsparadox. Die gaat erover dat rondom zo'n nieuwe technologie de verwachtingen vaak enorm hooggespannen zijn, maar dat de daadwerkelijke invloed op de productiviteit van de economie in ieder geval op korte termijn tegenvalt. Beroemd in deze zin is de opmerking van Robert Solow in 1987 dat de computer overal te vinden was, behalve in de productiviteitscijfers.

Ook dit is een vraagstuk dat in de context van AI is opgekomen. Het is het onderwerp van een artikel in de bundel van het National Bureau of Economic Research (NBER) over de *Economics of AI*, waarin de technologie als GPT behandeld wordt.[528] De auteurs constateren dat we, de hoge verwachtingen van AI ten spijt, een periode van zwakke productiviteitsgroei doormaken. Tussen 2005 en 2016 was de productiviteitsgroei in de VS slechts 1,3 procent per jaar, terwijl dat

526 Crawford et al. 2019: 14-16.
527 Das et al. 2020.
528 Brynjolfsson et al. 2019.

in de periode 1995 tot 2004 2,8 procent was. Studies van de OESO laten zien dat dit een breed verspreid mondiaal verschijnsel is. De auteurs concluderen ook dat deze vertraging niet aan de effecten van de mondiale recessie van 2008-2009 is toe te schrijven. Ze behandelen vervolgens drie redenen die dit fenomeen niet of in beperkte mate kunnen verklaren: valse hoop over de effecten van AI, slechte metingen van de productiviteitsgroei en een beperkte verspreiding van de voordelen van AI. Alleen voor de laatste verklaring is er in enige mate bewijs.

Vooral de verklaring van valse hoop over AI is belangrijk om onder de loep te nemen. Een uitgebreid argument daarvoor is namelijk ontwikkeld door Robert Gordon in zijn boek *The Rise and Fall of American Growth*.[529] In dat boek maakt hij ook vergelijkingen met eerdere grote technologische revoluties: de spoorwegen, het stoomschip en de telegraaf. Die eerdere technologieën zorgden voor immense verbeteringen in ons dagelijks leven: werk werd minder zwaar door mechanisering en huishoudelijke apparatuur, ziekte werd teruggedrongen door beter sanitair, elektrisch licht en ingeblikt voedsel maakten ons leven aangenamer en de levensverwachting maakte een enorme sprong. Volgens Gordon was zulke vooruitgang een eenmalige ontwikkeling; de huidige digitale technologieën zullen die niet herhalen. Hij laat dat zien aan de hand van productiviteitscijfers. Tussen 1920 en 1970 nam de productiviteit jaarlijks toe met 2,8 procent, om vervolgens – met een korte uitzondering tussen 1995 en 2005 – weer terug te zakken naar 1,7-1,8 procent, het niveau waarop ze ook vóór die periode lag. Dit cijfermatige verschil verklaart Gordon door op te merken dat digitale technologie vooral effect heeft op communicatie en minder dan die oudere technologieën op allerlei andere aspecten van het leven. Bovendien dragen hedendaagse ontwikkelingen rondom ongelijkheid en onderwijs ook bij aan een toekomstig lagere productiviteitsgroei.

Alhoewel Gordon een rijk betoog heeft, is het maar de vraag of hij voldoende rekenschap geeft van de recente doorbraken op het gebied van AI en of hij daarmee de potentiële effecten ervan niet onderschat. Veel van de hedendaagse verwachtingen gaan juist over sectoren buiten het gebied van communicatie, zoals mobiliteit, zorg en onderwijs. Zoals Carlota Perez betoogt, hoort het bij een volgende fase dat de effecten van een technologische revolutie door de hele economie heen verspreid worden, en daarmee de productiviteit doen stijgen.[530] Fenomenen waar Gordon op wijst zoals economische ongelijkheid kunnen zeker een negatief effect hebben op het breed verspreiden van de voordelen van AI, maar dat fenomeen is niet noodzakelijk met AI verbonden.

529 Gordon 2016.
530 Perez 2016.

De auteurs van de NBER-bundel stellen dan ook dat de productiviteitsparadox beter kan worden verklaard door de tijd die nodig is voor implementatie van en herstructurering als gevolg van de nieuwe technologie. De andere drie verklaringen gaan ervan uit dat één kant van de paradox niet klopt. Ze betogen ofwel dat de productiviteitsgroei er niet komt in het geval van valse verwachtingen (eerste verklaring) of ongelijke verdeling (derde verklaring) ofwel dat die groei er al wel is maar nog niet gemeten wordt (tweede verklaring). In de vierde verklaring, die uitgaat van vertraging, zijn beide observaties correct. De verwachtingen zijn terecht hoog, maar inderdaad vooralsnog niet gerealiseerd. Sterker nog, juist omdat de effecten zo'n grote verandering impliceren, is het logisch dat het tijd kost om die transitie te maken.[531] Dat is in lijn met ons idee van contextualisering. In het voorgaande benadrukten wij de technische vereisten die nodig zijn om AI te laten werken. Vanuit een macro-economisch perspectief gaat het daarbij om de ontwikkeling van nieuwe businessmodellen, het ontwerpen van allerlei andere processen in organisaties, efficiëntieslagen en prijsdalingen.

Zo kijkt de robotwetenschapper Rodney Brooks, die we eerder tegenkwamen bij de opgaven van demystificatie, ook naar AI. Hij gaat zo ver te stellen dat het dertig jaar duurt voordat iets dat in een laboratoriumdemonstratie werkt, ook een praktisch product kan worden. Zo ging het binnen AI bijvoorbeeld met het *backpropagation*-algoritme, waarbij technische doorbraken al in de jaren tachtig plaatsvonden.[532] Hetzelfde geldt voor de zelfrijdende auto. Zelfs als die technisch mogelijk is, blijven er vragen over hoe die is in te passen in de processen en regels van het verkeer: waar op de weg stopt de zelfrijdende auto om mensen op te halen? Hoe gaan andere weggebruikers reageren op deze voertuigen? Wat betekent het voor verkeerslichten en andere aspecten van de weg die afgestemd zijn op mensen en niet op autonome voertuigen?[533] In navolging van Solow zouden we kunnen stellen dat we de zelfrijdende auto op dit moment overal zien, behalve op de weg. Naast de techniek vraagt inbedding dus om een proces van maatschappelijke veranderingen en dat kost tijd. In een advies over robotisering spreekt de Sociaal-Economische Raad (SER) dan ook over de noodzaak van 'sociale innovatie'.[534]

531 Brynjolfsson et al. 2019: 42.
532 Ford 2018: 428-429.
533 Pasquale 2020: 21.
534 SER 2016.

Kernpunten – Het sociale ecosysteem: macro-economische context

- Ook rondom AI heerst de vrees dat de technologie tot massale werkloosheid zal leiden. Alhoewel de toekomst niet gekend kan worden, zijn er redenen voor twijfels over die vrees en spelen er urgentere vragen over het effect van AI op het werk.
- In plaats van netto banen te doen verdwijnen, lijkt AI vooral andere vaardigheden te vereisen van werkgevers en -nemers.
- Ook kan AI negatieve effecten hebben op de werkomstandigheden, bijvoorbeeld door surveillance van werknemers.
- Achter de manier waarop AI wordt ingezet en de effecten die ze heeft op werkgelegenheid, zitten economische en politieke keuzes.
- Naast de impact op werk speelt in de macro-economische context de vraag naar de productiviteitsparadox. Er zijn redenen om aan te nemen dat er een vertragingseffect is, juist omdat AI potentieel zoveel verandering in gang kan zetten.

De gedragsmatige context

Ook op het microniveau is het van belang aandacht te hebben voor de inbedding van AI in het sociale ecosysteem. Specifiek kijken we hier naar de gedragsmatige context waarin AI moet gaan werken. Een eerste observatie hierbij is dat vaak onvoldoende wordt nagedacht over hoe een nieuwe technologie een rol moet gaan spelen in de bestaande procedures en werkwijzen van mensen.[535] Dit speelt vaak op het gebied van de zorg. Ontwikkelaars komen met nieuwe software of een app, maar besteden geen aandacht aan de vraag hoe zorgprofessionals er gebruik van kunnen maken. Kunnen ze de app vertrouwen? Wie heeft toegang tot de data uit de app? Wat doet een dokter met een patiënt die thuis zelf een diagnose heeft gesteld met een app? In plaats van met een oplossing te komen voor een enkel aspect van het zorgproces, is het relevant om die technologie in te bedden in de bredere gedragspatronen van in dit geval zorgprofessionals en patiënten. Het Centrum voor Ethiek en Gezondheid (CEG) heeft studie gemaakt van de verschillende ethische uitdagingen die het gebruik van expertsystemen in de zorg met zich meebrengt.[536]

Een tweede hieraan gerelateerd punt is dat daarbij rekening gehouden moet worden met de ontvankelijkheid van mensen. Ook als iets goed werkt, kunnen mensen nog steeds redenen hebben om het te verwerpen. Een belangrijk aspect om rekening mee te houden is dat de technologie het werk van de betreffende

535 Omgaan met het conservatisme van mensen om nieuwe technologieën te gebruiken komt van wat Jane Bennett de 'materiële recalcitrantie van culturele objecten' noemt (Greenfield 2017: 307).

536 CEG 2018.

persoon in gevaar kan brengen. Ook dit is herkenbaar bij AI in de zorg. Wanneer ziekenhuizen of professionals beoordeeld worden op het aantal behandelingen, is een technologie die die behandelingen overbodig maakt een potentiële bedreiging. Om de nieuwe technologie te laten functioneren, kan het nodig zijn om een heel proces opnieuw in te richten zodat de drijfveren van mensen veranderen.[537] Ook in het onderwijs kunnen leraren AI als bedreiging zien. Een studie van Dialogic in opdracht van het ministerie van OCW benadrukt het belang van acceptatie, het stimuleren van digitale vaardigheden van onderwijzend personeel en het doen van experimenten.[538]

Een ander belangrijk gedragsmatig vraagstuk heeft te maken met het specifieke karakter van AI. In tegenstelling tot eerdere technologieën is het mogelijk dat AI in veel contexten zelf beslissingen gaat nemen die normaal gesproken mensen zelf nemen. Het vraagstuk is dan wat de juiste interactie moet zijn tussen mens en machine bij het nemen van bepaalde beslissingen.

Een aanvliegroute voor dat vraagstuk is een model dat drie vormen van mens-machine-interactie onderscheidt: *human-in-the-loop, human-on-the-loop* en *human-out-of-the-loop*. De eerste vorm van interactie is *human-in-the-loop*. Dit houdt in dat een AI-systeem in het proces betrokken kan zijn, maar dat de verantwoordelijkheid van de beslissing bij de mens ligt. Die is standaard onderdeel van de 'loop', waardoor zonder mens geen beslissingen mogelijk zijn. Een tweede vorm is *human-on-the-loop*. Hierbij heeft de mens een kleinere rol. In principe kan het AI-systeem hier zelfstandig beslissingen nemen zonder dat er een mens aan te pas komt. Wel heeft een mens inzicht in het proces en is die in staat om in te grijpen en wijzingen aan te brengen. Bij de laatste vorm, *human-out-of-the-loop*, is er geen mens meer bij het proces betrokken en handelt het AI-systeem volledig autonoom.

Er zijn veel situaties waarin het laatste model wordt toegepast. Dan gaat het om zaken die voor mensen niet van levensbelang zijn, zoals aanbevelingen voor films of producten. Het kan ook gaan om eenduidige situaties waarbij we erop vertrouwen dat algoritmes de juiste beslissingen nemen. Zo leiden de beelden van camera's tot automatische boetes voor snelheidsovertredingen zonder dat er een mens aan te pas komt.

---

537 Een interessant voorbeeld hiervan speelde eeuwen geleden met de introductie van koffie in Europa. In veel landen zagen de bierbrouwers koffie als een bedreiging en verzetten cafés zich tegen de opkomende koffiehuizen. Een oplossing in die tijd was om beide typen etablissementen zowel bier als koffie te laten schenken, waardoor ze geen concurrenten meer werden en de regering vervolgens gelijkmatig belasting kon gaan heffen op alle etablissementen (Juma 2016).

538 Van der Vorst et al. 2019.

Wanneer het situaties betreft die voor het leven van mensen van groot belang zijn, is het essentieel dat er wel een mens in het proces aanwezig zijn. De Europese privacywetgeving maakt dit een recht voor burgers. In artikel 15 van de EU Verordening Gegevensbescherming staat:

> "Member States shall grant the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc."[539]

Wat die beslissingen zijn die significant effect op burgers hebben, is niet precies gedefinieerd. Het afbakenen van die domeinen is een voortgaande discussie en de later gepubliceerde EU-aanbevelingen voor AI-gebruik bouwen daarop voort.[540]

In sommige domeinen kan het voldoende zijn om een mens *on the loop* te hebben om te controleren of er geen fouten worden gemaakt. Er zijn echter ook domeinen waar beslissingen zoveel invloed hebben dat menselijke controle, *in the loop* zijn dus, noodzakelijk wordt geacht. Dat geldt bijvoorbeeld voor zelfrijdende auto's. In tekstbox 5.6 lichten we dat verder toe. Bij militaire toepassingen speelt een situatie van leven en dood nog meer. Denk bijvoorbeeld aan autonome wapensystemen die eigenstandig hun doelwitten kunnen identificeren en uitschakelen. Verschillende legers experimenteren hier al uitgebreid mee, maar er is breed gedragen verzet tegen deze toepassing. Ook de Nederlandse Adviesraad Internationale Vraagstukken (AIV) heeft geadviseerd dat er bij autonome wapensystemen altijd 'betekenisvolle menselijke controle' moet zijn.[541]

Ook in andere contexten zoals fraudebestrijding of het toekennen van uitkeringen kunnen de effecten van beslissingen bijzonder ingrijpend zijn voor mensen, zoals de toeslagenaffaire laat zien. Dat pleit sterk voor blijvende menselijke controle in zulke gevallen. Hoe die controle invulling gegeven moet worden, is één van de uitdagingen bij het inbedden van AI in het sociale ecosysteem.

---

539 Algemene Verordening Gegevensbescherming, artikel 15.1.

540 De concept-Verordening AI (Europese Commissie 2021b) van de EU gaat uit van een risicobenadering ten aanzien van AI-systemen.

541 AIV en Commissie van Advies Inzake Volkenrechtelijke Vraagstukken 2015.

De drie vormen van mens-machine-interactie lijken een heldere manier om in verschillende contexten voor het juiste ontwerp te kiezen. Tegelijkertijd is er ook een aantal moeilijkheden bij deze benadering te identificeren.

Ten eerste kan het zo zijn dat mensen gedrag vertonen dat niet bij een gekozen model past. Bij mensen die officieel '*on*' of zelfs '*in*' de loop zijn, kan de aandacht bijvoorbeeld verslappen of ze kunnen zich roekeloos gedragen op basis van een ongerechtvaardigd vertrouwen in de technologie. *Automation bias* is een psychologisch mechanisme waarbij mensen blind de suggesties van een computer volgen, ook als die incorrect zijn. Een tegengesteld fenomeen is '*alert fatigue*'. Als een systeem teveel meldingen geeft en mensen met informatie overladen worden, nemen zij die meldingen minder serieus.[542]

Een tweede probleem is een sluipend proces waarin de menselijke beslissingsrol steeds minder betekenis krijgt. Denk aan een algoritme in de zorg dat ondersteuning biedt bij diagnoses. Doktoren nemen daarbij nog steeds de beslissingen en controleren de suggesties van de algoritmes. Lange gewenning aan het gebruik ervan kan die controle echter doen verslappen, zeker als de ervaring leert dat het algoritme goede diagnoses stelt. Op langere termijn kan het betekenen dat dokters aanvankelijk weliswaar de vaardigheden bezitten om zelf goede diagnoses te maken, maar dat nieuwe generaties artsen minder in die vaardigheid getraind worden. Dit komt overeen met het effect van rekenmachines op de vaardigheden van wiskundeleerlingen. Lange gewenning en het verhoogde tempo waarmee het werk gedaan kan worden, kunnen er bovendien voor zorgen dat het voor de menselijke beslisser moeilijker wordt om tegen de uitkomsten van een algoritme in te gaan. Die menselijke beslisser moet steeds sterker in de schoenen staan om een veelgebruikt en efficiënt proces te gaan betwijfelen.

Deze verschillende dynamieken maken de menselijke beslissing minder betekenisvol, terwijl mensen in deze situaties wel nog verantwoordelijk zijn voor de uitkomsten van die beslissing. Het risico ontstaat dan van een problematische fase waarin algoritmes nog niet goed genoeg zijn om beslissingen te nemen, terwijl mensen inmiddels niet meer in staat zijn om goed in te grijpen. Die situatie kan veel fouten, en daarmee menselijk leed, tot gevolg hebben.

In het verlengde daarvan ligt een derde uitdaging voor het interactiemodel. Menselijke controle is namelijk zinvol zolang algoritmes doen wat mensen normaal gesproken doen. Het is echter in veel contexten goed voorstelbaar dat met de tijd de activiteiten van het algoritme veel sneller en complexer worden.

542 Pasquale 2020: 37.

In die context is menselijke controle vaak niet meer mogelijk of zelfs gevaarlijk. Een mens moet nu bijvoorbeeld achter het stuur zitten van een autonoom voertuig om in te kunnen grijpen. Maar met de komst van C2C-communicatie (zie tekstbox 5.2) kunnen auto's veel dichter op elkaar gaan rijden. In dat geval is de menselijke reactiesnelheid te laag en wordt menselijke controle een gevaar op de weg. De opkomst van I2C-communicatie kan bovendien verkeersborden en zelfs verkeersregels overbodig maken, omdat voertuigen direct met hun omgeving communiceren. Verdwijnen die borden en regels echter, dan wordt het voor een mens heel moeilijk om op de weg navigeren. Bij de inzet van autonome wapens speelt eenzelfde moeilijkheid. Een mens kan een aanval op een individuele vijand overzien, maar wat wanneer het slagveld veel complexer wordt? Als het conflict bijvoorbeeld plaatsvindt tussen grote formaties van drones? Geen mens heeft het overzicht en de reactiesnelheid om daarop in te kunnen spelen.

John Danaher bespreekt een bestaand voorbeeld uit een heel ander domein. In klassieke magazijnen staan de producten geordend op categorieën. Door de legenda van die categorieën te kennen kan een mens de weg in zo'n warenhuis goed vinden. Bij de warenhuizen van Amazon bepaalt het *dynamic storage algorithm* wat de efficiëntste plaatsing van producten is. Allerlei producten staan daardoor kriskras door elkaar heen op basis van complexe berekeningen over toekomstige vraag en de logica daarvan is voor een mens niet te doorgronden. Alleen door een algoritme aan te sturen, kan een mens nog de weg vinden in zo'n warenhuis. Danaher spreekt in dit verband van het gevaar van 'algocratie', een bestel dat terugvoert op complexe algoritmes en onnavolgbaar is voor mensen.[543]

Tekstbox 5.6 – De gedragscontext van autonoom vervoer

De verschillende vraagstukken rondom de gedragscontext van AI spelen bij uitstek bij autonoom vervoer. Wettelijk zijn er nog geen zelfrijdende auto's toegestaan. Dat heeft niets te maken met de technische mogelijkheden, maar met het feit dat op dit moment menselijke bestuurders de verantwoordelijkheid moeten dragen voor het bestuur van alle auto's op de Nederlandse weg. Tegelijkertijd gedragen mensen zich daar vaak niet naar, met serieuze gevolgen.

543 Danaher 2016.

Op een basaal niveau speelt dit al bij navigatiesoftware. Tegen gezond verstand en actuele verkeerstekens in neigen mensen er soms toch naar de instructies van het navigatiesysteem te volgen. In klassieke voorbeelden van *automation bias* zijn mensen bijvoorbeeld de zee of onbegaanbare natuurpaden in gereden. Fatale incidenten hiermee staan bekend als 'death by GPS' (James Bridle 2018).

Hoewel veel aandacht uitgaat naar zelfrijdende auto's, zijn die dus nog geen realiteit. Ondertussen zijn er wel allerlei beslissingsondersteunende programma's zoals ADAS, waarbij mensen weliswaar verantwoordelijk zijn voor de beslissing maar daar onvoldoende naar handelen, met ongelukken tot gevolg. Dat is een van de conclusies van een rapport van de Onderzoeksraad voor Veiligheid. De menselijke factor moet bij de automatisering van vervoer dan ook serieuzer worden genomen bij het inschatten van risico's.

Heel specifiek speelt dit bij een functie van de auto's van Tesla, genaamd 'autopilot'. De naam suggereert dat de bestuurder achterover kan leunen, maar, zoals in de gebruiksaanwijzing staat, dat is niet het geval. Ondanks kritiek op de suggestie die ze wekt, houdt het bedrijf vast aan deze misleidende benaming. Goede communicatie en instructie zijn dus zeer relevant vanuit het oogpunt van menselijk gedrag.

Daaraan gerelateerd is het effect van updates op gedrag. Updates zijn bedoeld om de auto's beter te maken. Tegelijkertijd kunnen zij er wel voor zorgen dat een auto op één moment op een bepaalde manier op een specifieke situatie reageert, en op een later moment anders reageert in precies dezelfde situatie. Dat kan moeilijkheden en verwarring voor de bestuurder veroorzaken. Ook ergonomische zaken, zoals het duidelijk aangeven aan de bestuurder wat voor functie in de auto nu aan staat en heldere afbakening van andere functies, zijn van belang om gevaren door menselijk gedrag te voorkomen.

Het bovengenoemde risico van een riskante tussenfase speelt hier ook. Terwijl auto's nog niet in staat zijn om autonoom alle beslissingen op de weg te nemen, kan van mensen niet verwacht worden dat ze aandachtig bij de weg blijven tijdens lange ritten waarop de auto stuurt. Sommigen pleiten er dan ook voor om vanwege deze menselijke factor niet te experimenteren met semi-autonome voertuigen: óf de mens rijdt óf de auto, maar niets daartussenin.

De genoemde drie uitdagingen plaatsen vraagtekens bij een te gemakkelijke nadruk op menselijke controle in veel contexten. Ze vragen om aandacht voor de complexiteit van vraagstukken rondom mens-machine-interactie. Die complexiteit gaat ook over het identificeren van de sterke en zwakke punten van mensen en machines, die niet hetzelfde zijn. Waar een machine bijvoorbeeld veel beter patronen in grote hoeveelheden data kan ontdekken, kan een mens vaak beter redeneren over anomalieën. Een goede interactie vergt de juiste afstemming tussen de eigenschappen van mens en machine, waarbij ze elkaars zwakten compenseren en elkaars krachten tot hun recht laten komen. Catholijn Jonker spreekt in deze zin van '*hybrid intelligence*'.[544]

De eigenschappen van AI-systemen kunnen verschillen, afhankelijk van hoe ze afgesteld zijn. Een AI-systeem dat menselijke redacteuren ondersteunt bij het zwartmaken van passages in een tekst, kan op verschillende manieren opgezet worden. Als het risico vooral is dat gevoelige informatie toch naar buiten komt, kan het algoritme streng afgesteld worden. Als de vrees vooral is dat te weinig openbaar wordt gemaakt, kan zo'n algoritme veel lichter worden afgesteld.[545] Hoe het systeem afgesteld dient te worden, hangt dus af van de context waarin het moet opereren.

Kernpunten – Het sociale ecosysteem: gedragscontext

- Inbedding in de gedragsmatige context betekent rekening houden met bestaande organisatiestructuren, werkwijzen en motieven van menselijk gedrag.
- Het model van *human in/on/out of the loop* is een manier om de interactie tussen mensen en machines in te richten en onderscheid te maken tussen verschillende maten van menselijke controle.
- Allerlei gedragsfactoren kunnen deze sterke onderscheiding echter ondermijnen en vragen daarom om uitgebreide aandacht in het ontwerp en gebruik van technologie.

544 TU Delft z.d.
545 Agrawal et al. 2018: 68.

5.3 Tot slot

In dit hoofdstuk hebben we aangegeven wat de stand van zaken in Nederland is met betrekking tot de opgave van contextualisering in het socio-technische ecosysteem. Een groot deel van dit proces kan niet centraal gestuurd worden. Het zal plaatsvinden in allerlei organisaties die experimenteren met hun processen of die door betere productiemethoden efficiëntieslagen realiseren. De AI-Coalitie werkt in dit kader aan het opzetten van sectorale databases.

De overheid kan hier desalniettemin een belangrijke rol bij spelen. Ten eerste door investeringen, bijvoorbeeld in goede digitale infrastructuur of bijscholing. Daarnaast kan ze contextualisering beïnvloeden door het eigen gebruik van AI. Overheidsorganisaties, vooral uitvoeringsorganisaties, kunnen goede praktijken van contextualisering helpen ontwikkelen en zelfs standaarden stellen. Ook kan de overheid bijdragen via haar aanbestedingsbeleid. Als grote speler kan het zo als *launching customer* beginnende markten stimuleren. De Raad voor de leefomgeving en infrastructuur (Rli) ziet op die manier een rol voor de overheid om digitale technologie in te zetten voor duurzaamheid[546], en de Cybersecurity Raad ziet een rol voor haar om actief capaciteiten te ontwikkelen op het gebied van cybersecurity.[547]

In hoofdstuk 8 over positionering komen we terug op vraagstukken rondom het verdienvermogen van Nederland. Hier is het van belang te noemen dat de overheid met haar brede palet aan instrumenten de mogelijkheid heeft om prioriteiten te stellen voor domeinen waar AI toegepast gaat worden en om daar contextualisering te stimuleren. Het kan dan gaan om domeinen die met verdienvermogen en de economische motor van ons land samenhangen, zoals land- en tuinbouw of infrastructuur, maar ook om domeinen die maatschappelijk van groot belang zijn. Denk in die categorie aan terreinen waar de overheid een specifieke verantwoordelijkheid heeft om een voortrekkersrol te vervullen, zoals de zorg of de verduurzaming. Op die manier kan Nederland meer inzetten op een eigen 'AI-identiteit'.

546 Rli 2021.
547 Cybersecurity Raad 2020.

6. Engagement

De volgende opgave voor de maatschappelijke inbedding van AI die we onderscheiden, is het engageren van belanghebbenden. De bijbehorende vraag van deze opgave is: *Wie moeten er betrokken zijn?* Bij de introductie van elke technologie zijn immers vanaf het begin al verschillende partijen betrokken. Dat bleek ook in het vorige hoofdstuk over contextualisering. Daar zagen we dat zowel bedrijven als de overheid al vroeg met AI aan de slag zijn gegaan. Het perspectief van waaruit we die partijen toen hebben besproken, was de vraag: hoe laten we AI functioneren? Vanwege hun positie en middelen zijn bedrijven en overheidsorganisaties belangrijke aanjagers van het gebruik van AI in de samenleving. Hierdoor hebben ze ook veel invloed op hoe AI in de praktijk wordt gebracht. In dit hoofdstuk gaat de aandacht expliciet uit naar de partijen die in eerste instantie niet zelf met AI aan de slag gaan, maar daarmee wel te maken kunnen krijgen, gezien het alomtegenwoordige gebruik ervan. We richten ons in het bijzonder op de partijen uit het maatschappelijk middenveld.

Juist omdat bepaalde groepen meer capaciteit hebben om een nieuwe systeemtechnologie te gebruiken dan andere groepen, zo zagen we in hoofdstuk 3, gaat de introductie van zo'n nieuwe technologie gepaard met maatschappelijke spanningen en groeiende ongelijkheid. Werknemers kwamen tijdens de Industriële Revolutie in precaire situaties terecht. Het platteland bleef lange tijd achter met de elektrificatie van de samenleving, en de auto werd geassocieerd met de rijke witte bevolking en joeg armere weggebruikers opzij. Er was ook meer indirect leed. Elektrisch licht werd gebruikt om werknemers beter te controleren en auto's vervuilden de luchtkwaliteit en maakten fietsen op de weg gevaarlijk. Het proces van inbedding ging dus vrijwel altijd gepaard met misstanden en onverantwoord gebruik van de nieuwe technologie. We zagen ook dat groepen in het maatschappelijk middenveld zich mettertijd mobiliseerden om die misstanden tegen te gaan en de disbalans in het gebruik van de nieuwe technologie te corrigeren. Bij de introductie van een nieuwe systeemtechnologie is het daarom belangrijk om verschillende groepen bij dit proces te betrekken om zo de nieuwe technologie mede vorm te kunnen geven.

De betrokkenheid van het maatschappelijk middenveld bij de inbedding van AI volgt om te beginnen uit het gegeven dat elke samenleving uit een veelheid van verschillende partijen bestaat. Alleen de meest autoritaire regimes zullen een enkele speler – een leider of politieke partij – aanwijzen die de richting van de samenleving bepaalt. In een democratische rechtsstaat als Nederland bestaat een pluraliteit aan instituties die voor een balans met de staatsmacht kunnen zorgen, maar tegelijk ook door die staat worden beschermd, onder andere door constitutionele rechten. Een sterk en goed ontwikkeld maatschappelijk

middenveld is dan ook een belangrijke maatschappelijke voorwaarde voor het goed functioneren van staat en markt.[548] Het maatschappelijk middenveld kan gebruik maken van een groot aantal manieren om aan te haken bij de inbedding van een nieuwe systeemtechnologie. Deze manieren moesten tijdens de Industriële Revolutie in hoge mate nog worden uitgevonden, omdat arbeiders niet waren verenigd en algemeen stemrecht ontbrak. Tegenwoordig kunnen belanghebbenden veel gemakkelijker dan voorheen hun stem laten horen, bijvoorbeeld door rechtszaken aan te spannen, nieuwe belangenorganisaties op te richten en inspraak te hebben bij de besluitvorming in zowel publieke als private organisaties. Het betrekken van belanghebbenden is zeker niet altijd vrijblijvend. Illustratief is het wettelijk verankerde instemmingsrecht van de Ondernemingsraad (neergelegd in de Wet op de Ondernemingsraden) wanneer een onderneming personeelsvolgsystemen wil gebruiken, zoals camera's op de werkvloer of intelligente trackingsystemen in vrachtwagens. Het engagement van belanghebbenden is dus behalve een zaak van particulier initiatief ook een formele verplichting.

Dat belanghebbenden zich kunnen engageren met zaken die impact hebben op hun leven, is in een democratische samenleving een waarde op zichzelf. Daarnaast kan een brede betrokkenheid van de samenleving een technologie ook beter maken.[549] Partijen reageren niet alleen op de effecten van het gebruik van AI, maar brengen ook hun kennis en ervaring in en kunnen de technologie zelfs gaan gebruiken om hun eigen belangen en waarden te behartigen. Al vanaf de jaren zestig zijn er initiatieven om belanghebbenden te betrekken bij technologieontwikkeling, om zicht te hebben op impact daarvan op de samenleving en om technologiegebruik maatschappelijk meer te verantwoorden. Door waarden en morele overwegingen mee te nemen bij de ontwikkeling van een nieuwe technologie, kan bovendien worden voorkomen dat de toepassing ervan in een later stadium alsnog op maatschappelijk verzet stuit.[550]

Net als bij eerdere systeemtechnologieën ontstaan rondom AI allerlei misstanden en worden machtsongelijkheden versterkt, zo zullen we in dit hoofdstuk zien. Door deze problemen zichtbaar te maken en aan te kaarten, draagt het maatschappelijk middenveld wezenlijk bij aan de verdere inbedding van AI in de gehele samenleving. Engagement is, met andere woorden, een belangrijke opgave voor de inbedding van AI in de samenleving. Verschillende partijen uit

548 Schuyt 2006.
549 Sikes en Macnaghten 2013: 85-107.
550 Nederlandse voorbeelden zijn het elektronische patiëntendossier (EPD) en de slimme energiemeter. Zie hiervoor Van den Hoven 2013: 75-83.

het maatschappelijk middenveld zijn voor deze opgave in het bijzonder van belang, zoals belangengroepen, de media, wetenschappers en andere experts.

Centraal bij de effecten van een nieuwe technologie op benadeling en gelijkheid staat het vraagstuk van de expertise van belangengroepen. Hoewel er in Nederland een veelheid aan organisaties is die zich bezighoudt met het helpen van benadeelde groepen in de samenleving, zijn veel van die organisaties niet voldoende toegerust om dat werk te blijven doen in het licht van een nieuwe technologie als AI. De onterechte benadeling van bepaalde groepen krijgt door AI namelijk nieuwe dimensies waardoor de aard van het werkveld van die organisaties en de problemen die zij moeten adresseren verandert. We zouden kunnen zeggen dat er AI-varianten ontstaan van verschillende vormen van ongelijkheid die vragen om een andere blik en aanvullende expertise.

Dat geldt bijvoorbeeld voor discriminatie tegen mensen van kleur. Ruha Benjamin muntte daarvoor de term de New Jim Code. Die term verwijst naar de Jim Crow-wetten die raciale segregatie in het zuiden van de VS codificeerden. De hedendaagse code benadeelt ook raciale minderheden, maar op een andere manier. Benjamin definieert de New Jim Code als "het gebruik van nieuwe technologieën die bestaande ongelijkheden reflecteren en reproduceren, maar die gepromoot en waargenomen worden als objectiever en progressiever dan de discriminatoire systemen van een eerder tijdperk."[551] Niet alleen kan discriminatie dus via AI plaatsvinden, dat gebeurt bovendien op een manier die veel subtieler is.[552] Waar er de laatste tijd bijvoorbeeld veel aandacht uitgaat naar discriminatie door politieagenten, gaat het hier om een andere vorm van discriminatie, die als objectief wordt gepresenteerd en minder zichtbaar is. Dat maakt hem ook moeilijker te identificeren en te bestrijden; hier is namelijk geen racistische baas, bankier of winkeleigenaar om aan te geven.[553] Sterker nog, het principe van discriminatie wordt vaak als iets positiefs gepresenteerd. Een dienst als Netflix maakt bijvoorbeeld verschillende trailers voor verschillende doelgroepen. Dat kan betekenen dat iemand die dat belangrijk vindt, een trailer te zien krijgt waarin vooral acteurs van kleur te zien zijn. Daarmee wordt dat de suggestie gewekt dat series diverser zijn dan ze in werkelijkheid zijn. Terwijl bij de uitreiking van de Oscars iedereen kan zien hoe divers de winnaars zijn, maakt Netflix dat moeilijk omdat iedereen een andere representatie krijgt van acteurs en producenten. Safiya Umoja Noble spreekt in dit verband ook wel van "algoritmes van onderdrukking".[554]

551 Benjamin 2019: 5.
552 Zie ook: Wallace 2021.
553 Benjamin 2019: 33.
554 Noble 2018.

Of neem uitsluiting van mensen met lage inkomens. Virginia Eubanks beschrijft hoe vroeger in de VS arme mensen onderdrukt en gestigmatiseerd werden in het 'poorhouse'. Ze werden naar dit armenhuis gestuurd omdat ze lui zouden zijn en moesten daarvoor vaak onbetaald arbeid verrichten. In plaats van dit armenhuis bestaat er volgens Eubanks nu een *'digital poorhouse'*.[555] Datapunten van burgers stigmatiseren hen waardoor het moeilijker kan zijn om verzekeringen, hypotheken en uitkeringen te krijgen.[556] Eubanks documenteert hoe dit digitale armenhuis op allerlei plaatsen op subtiele wijze wordt toegepast. Ook hier geldt dat het maken van onderscheid zelfs expliciet als iets positief wordt gepresenteerd. Illustratief is het groeiend aanbod aan verzekeringsproducten waarbij korting wordt verleend in ruil voor persoonsgegevens. Aan de hand van die gegevens kunnen deze bedrijven het gedrag van verzekerden beter voorspellen en hen zo een passender aanbod doen. Transparantie over de gehanteerde criteria en foutmarges, parameters en analytische inzichten bieden de betreffende partijen bij deze vormen van *prijsprofilering* zelden.[557] En met de sluipende acceptatie van deze praktijk wordt op subtiele wijze ongelijkheid in de hand gewerkt.

Een volgende kwestie is de schending van mensenrechten. Een traditionele vorm daarvan is het in gevangenissen opsluiten van activisten en mensen met afwijkende meningen. Technologie als AI maakt het tegenwoordig mogelijk om afsluiting en opsluiting op digitale wijze te organiseren. Denk aan het ontwikkelde sociale kredietsysteem in China dat mensen met een lage score uitsluit van treinen en vliegtuigen. Mensenrechtenorganisaties spreken ook van een 'open air prison' in de Chinese provincie Xinjiang. In hoofdstuk 8 gaan we daar verder op in.

Een laatste voorbeeld van een AI-variant van ongelijkheid heeft betrekking op gender en seksuele geaardheid. Caroline Criado Perez laat zien hoe in veel domeinen van werk tot zorg en politiek de standpunten en belangen van mannen centraal staan. Bij de digitalisering van die domeinen zijn data van vrouwen dan ook ondervertegenwoordigd. Vrouwen worden beschouwd als 'kleinere mannen', in plaats van te erkennen hoe vrouwen ook kunnen verschillen van mannen. Veel AI-toepassingen werken dan ook niet goed voor vrouwen.[558] Ook uitsluiting van mensen met een andere seksuele geaardheid krijgt een nieuwe vorm. Frank Pasquale laat bijvoorbeeld zien dat in de wereld van kredietscores

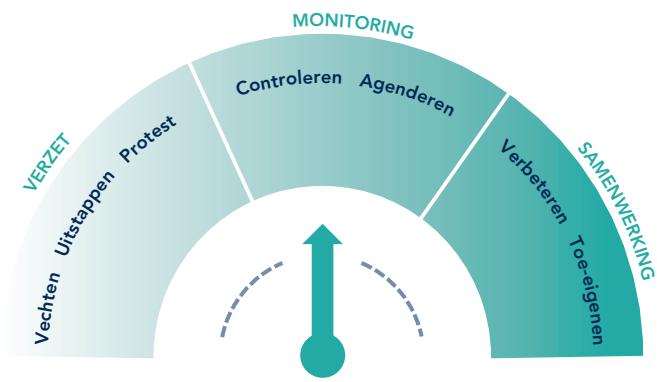
555 Eubanks 2018.
556 Om deze reden besteden AFM en DNB in een verkenning over AI-gebruik in de verzekeringssector ruim aandacht aan het risico van discriminatie, vooral bij modellen voor beprijzing, acceptatie en fraudedetectie. Zie: AFM en DNB 2019.
557 Moerel en Prins 2016.
558 Perez 2019.

de aanwezigheid van homoseksuele mannen juist als een positieve indicator wordt gezien voor huizenprijzen.[559] In tegenstelling tot het verleden wordt homoseksualiteit hier dus niet als iets negatiefs bestreden. Tegelijkertijd gaat het hier wel een vorm van andere behandeling die op subtiele wijze ongelijkheid in de hand kan werken.

We zien dus een patroon op alle bovenstaande gebieden. In de geschiedenis is een aantal groepen als afwijkend bestempeld en die groepen werden gedisciplineerd om te voldoen aan een maatschappelijke norm. In een wereld van AI wordt verschil niet op die manier bestreden, maar fungeert deze technologie indirect als bron van ongelijke behandeling, vaak zelfs terwijl dat verschil – mede onder het mom van kansen voor een persoonsgerichtere benadering – als iets positiefs wordt gepresenteerd.

In dit hoofdstuk bespreken wij deze opgave van engagement aan de hand van verschillende vormen die betrokkenheid kan aannemen: vechten, uitstappen, protest, controleren, agenderen, verbeteren en toe-eigenen. We plaatsen die vormen op een continuüm die de verhouding tot AI karakteriseert (figuur 6.1). In het ene uiterste is de houding antagonistisch: groepen bestrijden AI of willen de technologie bijvoorbeeld verbieden. Dit cluster bespreken we onder de noemer 'verzet'. In het andere uiterste is de houding symbiotisch: groepen engageren zich met AI door de technologie in hun dagelijks leven op te nemen. Deze houding vatten we samen als 'samenwerking'. Hiertussen in neemt de betrokkenheid een kritische vorm aan, die we duiden met 'monitoring'.

Figuur 6.1 Een spectrum van verschillende vormen van engagement



559 Pasquale 2015: 25.

Deze verschillende vormen van engagement zijn ideaaltypisch en lopen in de praktijk vaak in elkaar over. We zullen ook zien dat bij verschillende partijen meerdere vormen van engagement naast elkaar bestaan. Het geschetste spectrum is een manier om overzicht te krijgen over dit brede veld van activiteiten die allerlei partijen in het maatschappelijk middenveld ontplooien, met als doel om bestaande praktijken, en soms ook geldende wetten en regels, op te rekken, te buigen, te breken, of te verleggen. We bespreken hieronder de verschillende vormen van engagement en laten daarbij zien welke status die vormen hebben ten aanzien van AI. Daarmee wordt duidelijk welke vormen prevaleren en waar nog meer werk te doen is.

6.1 Verzet

Een nieuwe technologie roept vaak verzet op. Zeker wanneer technologische verandering snel gaat en het idee postvat dat de baten van de technologie slechts ten goede komen aan een zeer beperkt deel van de samenleving, terwijl de risico's ervan wijdverspreid zijn.[560] Verzet plaatsen we links in het spectrum van verschillende vormen van engagement en dit omvat drie verschillende uitingen. Uiterst links staat de meest antagonistische verhouding tot AI: vechten. Belanghebbenden wijzen de nieuwe technologie af en gebruiken daar ook geweld bij. Na vechten volgt 'uitstappen'. Hierbij gaat het om de handelingsoptie die Albert Hirschman 'exit' noemde. Wanneer belanghebbenden gebruik maken van exit, geven zij hierover een signaal door van de onderhandelingstafel weg te lopen. Hirschman plaatst 'exit' tegenover 'voice'. Bij dat laatste articuleren belanghebbenden hun onvrede zonder de relatie te verbreken. Een voorbeeld daarvan is de derde vorm van verzet die we onderscheiden: protesteren. Ook hier wordt de technologie bestreden, maar op vreedzame wijze en met een duidelijk gearticuleerd tegenvoorstel, bijvoorbeeld een verbod van een technologie of een nadere regulering van de toepassing daarvan.

Vechten: gewelddadig verzet

In de geschiedenis is gewelddadig verzet vaak een iconische vorm van negatief engagement bij een nieuwe technologie. In hoofdstuk 3 noemden we al de befaamde Luddieten, die bezorgd waren over het verdwijnen van hun banen en inkomens en er daarom toe overgingen om de nieuwe machines kapot te slaan. Zij hadden ook niet veel andere mogelijkheden voor hun verzet. Naar hun stem luisterden de eigenaars van de machines niet en ook politiek werden zij niet vertegenwoordigd.

In moderne democratieën hebben groepen die door een technologie geschaad worden, allerlei niet-gewelddadige mogelijkheden om hun stem te laten horen.

560 Juma 2016

In hoofdstuk 3 hebben we ook betoogd dat democratisering de inbedding van latere systeemtechnologieën onderscheidt van die van eerdere. Toch vindt ook in democratische samenlevingen verzet plaats dat doelbewust wettelijke grenzen overschrijdt en geweld daarbij niet schuwt. Denk aan de antikernwapenbeweging die centrales binnendrong en materieel vernielde.[561]

Wordt er op dit moment ook tegen AI gevochten? Vooralsnog lijkt dat niet het geval. Dat kan ermee te maken hebben dat zowel de technologie als haar effecten minder direct zichtbaar zijn. Daardoor is ze moeilijker te lokaliseren en vernielen. Vanwege het immateriële karakter kan weerstand tegen AI ook vermengd raken met weerstand tegen andere fysiekere zaken als computers, robots of de bedrijven die AI ontwikkelen.

In 2014 keerde een anarchistische beweging, The Counterforce, zich bijvoorbeeld tegen de invloed van Silicon Valley en in het bijzonder tegen die van bedrijven als Google. De aanklacht van de groep was breed en gericht op het effect dat deze bedrijven hadden op de huizenprijzen in San Francisco, en daarmee op de levenskwaliteit van gewone mensen. Ze ageerden ook tegen het effect van digitale technologie op de aandacht van mensen en de bouw van een infrastructuur die totalitair gebruikt kan worden. Het verzet van The Counterforce ging verder dan demonstreren. The Counterforce moedigde het blokkeren van bussen van medewerkers in Silicon Valley aan, maar ook het stelen van bezittingen van techneuten en het neerhalen van surveillancecamera's. Een van de mensen waar deze beweging zich tegen richtte, was Anthony Levandowski, die op dat moment verantwoordelijk was voor de technologie achter Googles zelfrijdende auto.[562] In Hongkong gingen protesterende mensen lantaarnpalen die waren uitgerust met gezichtsherkenning, te lijf met een elektrische zaag.[563]

Dichter bij huis is AI wel degelijk onderdeel van zaken waar groepen zich met geweld tegen verzetten. Dat gebeurt in het narratief op sociale media rondom 5G en coronavaccins. Amerikaanse filmpjes op YouTube verbinden 5G, Huawei en AI met Chinese plannen om in het geniep wereldwijd data te verzamelen. Er zijn ook mensen die dachten dat COVID-19 stond voor "certificate of vaccination identification by AI", waarbij AI als eerste en negende letter van het alfabet voor 19 in COVID-19 zouden staan.[564] Deze theorie komt van een ostheopate met de naam Carrie Madej en circuleerde in het voorjaar van 2020 op het internet. Het coronavaccin zou ons DNA herschrijven zodat we allemaal in een interface

561 Van der Vleuten et al. 2017: 135
562 Jeffries, 15 april 2014.
563 Fussel, 30 augustus 2019.
564 Reuters, 24 april 2020.

(API) van mens en machine worden geplaatst waardoor ons gedrag volledig van buitenaf te sturen zou zijn. Bill Gates zou hierachter zitten. In dergelijke samenzweringstheorieën is AI onderdeel van een narratief dat mensen ertoe gebracht heeft telecomzendmasten te vernielen. Deze problematische vorm van engagement behoeft geen versterking en er dient juist gewaakt te worden voor escalatie. De eerder besproken opgave van demystificatie kan ook bijdragen aan meer vreedzame en democratische betrokkenheid bij AI.

Uitstappen: weigeren medewerking te verlenen

Verzet neemt een geweldloze vorm aan wanneer mensen hun medewerking ergens aan weigeren. Deze tweede vorm engagement die we hier onderscheiden, noemen we uitstappen. Het gaat hierbij onder meer om mensen die werkzaam zijn in de technologiesector. Zij hebben een bijzonder instrument tot hun beschikking door hun werk te staken. Individuen met bijzondere expertise kunnen druk uitoefenen door te weigeren ergens aan mee te werken. Zonder hun inbreng kunnen bepaalde projecten niet van start gaan. Mensen met minder schaarse expertise kunnen collectief druk uitoefenen, in het bijzonder wanneer zij aandacht voor hun actie weten te genereren. De laatste jaren groeit deze vorm van engagement. Er waren bijvoorbeeld verschillende acties van dit type in Silicon Valley, de zogenoemde 'walkouts'. Maar ook Nederland kent deze vorm van engagement. Illustratief is de juridische procedure tussen de Universiteit van Amsterdam en studenten van deze instelling. Studenten weigerden zich tijdens tentamens te laten observeren door op AI gebaseerde onlinesurveillancesoftware (proctoring) en kozen in die zin ook voor een 'walkout'.[565] De weigering, bijna tien jaar geleden, van diverse burgers om hun vingerafdruk op te laten nemen in het paspoort, toen daarvoor biometrie werd geïntroduceerd, is een voorbeeld van recenter datum. De actie was destijds aanleiding voor onder meer de Nationale Ombudsman om bij toenmalig minister Plasterk van Binnenlandse Zaken te pleiten voor een afzonderlijke regeling voor deze weigeraars.[566]

Veel walkouts zijn in feite werkonderbrekingen en betreffen logischerwijs de werkomstandigheden. Alhoewel acties om betere werkomstandigheden bij alle soorten bedrijven kunnen voorkomen, zijn deze acties soms ook verbonden met het gebruik van technologieën als AI. Een nieuwe systeemtechnologie creëert namelijk ook nieuwe werkomstandigheden die voor werknemers heel nadelig kunnen zijn. Net als in het verleden de introductie van elektrisch licht

565 In de rechtszaak die hierover werd aangespannen, oordeelden overigens zowel de rechtbank als het hof dat het gebruik van de software door de UvA rechtmatig was (Rechtbank Amsterdam, 11 juni 2020; Gerechtshof Amsterdam, 1 juni 2021).

566 Blankena, 9 september 2013.

de surveillance van werknemers versterkte, doen algoritmes dat ook. Dat gaat van het monitoren van het surfgedrag van kantoorwerknemers, en zelfs hun biometrische informatie zoals oogbewegingen, tot het minutieus aansturen van personeel in magazijnen en bezorging (zie tekstbox 6.1).[567]

Tekstbox 6.1 – Surveillance werknemers

Het vergaren van data over werknemers en ze daarmee aansturen gebeurt al zeker honderd jaar. Doordat de technologie is verbeterd, is nu meer diepgaande, variabele, fijnmazige, grootschalige en snelle surveillance en monitoring van werkzaamheden mogelijk.[568] Het gebruik van AI bij de analyse van werknemersdata vindt veelal plaats in het kader van personeelsbeleid.[569] Daarbij kunnen op grond van de gebruikte data en de beoogde doeleinden vier deeltrends onderscheiden worden:

1. Systemen die diverse datasoorten gebruiken om voorspellingen over (mogelijk ongewenst) gedrag van werknemers doen;
2. Systemen die op basis van biometrie en gezondheidsdata inferenties over arbeidsomstandigheden maken, werknemers inzicht in hun gezondheid geven, maar ook als volgsystemen dienen;
3. Systemen die het gedrag van werknemers op afstand monitoren om hun prestatie vast te stellen en hun beloning te bepalen;
4. Systemen die middels voortdurende vergaring van prestatiedata 'algoritmische aansturing' of 'gamificatie' van het werk mogelijk maken.[570]

Ook in Nederland zijn dergelijke systemen zichtbaar. Medewerkers van PostNL worden bijvoorbeeld door een app aangestuurd, die de postbode de te volgen route en de daarvoor beschikbare tijd voorrekent. Het overschrijden van de beschikbare tijd heeft gevolgen voor de medewerkers, maar de app houdt weinig rekening met zaken als het weer of overgebleven post van de vorige dag.[571] Ook bij veel andere bedrijven en overheden verzamelen werkgevers, vaak zonder dat hun personeel het weet, data over onder meer hun gemoedstoestand, gezondheid en zelfs bevlogenheid om het personeelsbeleid meer op feiten te kunnen baseren. Ze gebruiken daartoe vaak taalanalyse.

567 Het AI Now Report van 2019 beschrijft bijvoorbeeld het effect van het gebruik van 'the rate' om werknemers aan te sturen in de warenhuizen van Amazon.
568 Mateescu en Nguyen 2019.
569 Zie onder andere Das et al. 2020.
570 Mateescu en Nguyen 2019.
571 Kuijper et al., 31 oktober 2018.

De e-mails van (groepen) werknemers worden op zaken als 'bevlogenheid' geanalyseerd, het taalgebruik zou daar indicaties van bevatten. Het probleem met dit soort analyses is echter dat de validiteit ervan niet bewezen is.

Een nieuw type werkomstandigheid dat door nieuwe technologie is ontstaan, is het platformwerk.[572] Chauffeurs van taxi's of bezorgers van maaltijden of pakketten zijn geen officiële werknemers met rechten op minimumloon of bepaalde secundaire arbeidsvoorwaarden, waardoor voor velen het levensonderhoud precair is. De opkomst van vakbonden, maar ook rechtszaken om deze mensen wel als werknemers te erkennen zijn hiertegen gericht. In ons land bevestigde het gerechtshof Amsterdam in februari 2021 een eerdere uitspraak van de rechtbank dat bezorgers van Deliveroo een arbeidsovereenkomst hebben.[573]

De VS kende verschillende grote acties van werknemers specifiek gericht op AI. In 2018 gaf een groep ingenieurs bij Google aan niet te willen meewerken aan Project Maven. Dit project van het Amerikaanse leger is erop gericht om drones te voorzien van geavanceerde beeldherkenning zodat zij mensen en objecten automatisch kunnen herkennen. Als gevolg van de grote onvrede bij de werknemers continueerde Google dit project met defensie niet. Later dat jaar tekenden Google-ingenieurs een petitie tegen Dragonfly, een door het bedrijf ontwikkelde gecensureerde zoekmachine voor China waarmee Google in dat land voet aan de grond wilde krijgen. De werknemers weigerden bij te dragen aan onderdrukking en ook dit project werd later gestaakt. Microsoftmedewerkers hebben zich in 2019 publiekelijk via een open brief verzet tegen het inschrijven op de aanbestedingen van JEDI, een cloud-computing project van het Amerikaanse leger, en augmentedrealityapparatuur, omdat zij niet van oorlog wilden profiteren.

Een ander toepassingsgebied van AI waar werknemers zich tegen verzetten, is het leveren van technologie aan de Amerikaanse grensbewakingsinstantie Immigration and Customs Enforcement (ICE). Werknemers van Palantir, Salesforce, Microsoft, Accenture, Google, GitHub en Tableau tekenden in 2018 petities en open brieven tegen werkzaamheden voor die organisatie. In een brief gepubliceerd in *The New York Times* riepen Microsoft-werknemers hun CEO Satya Nadella op om "een ethische positie in te nemen en kinderen en families

572 Frenken en Fuenfschillling 2020.
573 Gerechtshof Amsterdam, 16 februari 2021.

boven winst te plaatsen"[574]. Als gevolg van deze acties sprak de CEO zich uit tegen het migratiebeleid van president Trump. Het bedrijf Chef Robotics werkte eveneens voor ICE. De programmeur Seth Vargo verwijderde de door hem geprogrammeerde codes die het bedrijf gebruikte, waardoor de dienst een aantal dagen gestaakt moest worden. Uiteindelijk stopte Chef Robotics de samenwerking met ICE.

Ook op individueel niveau kunnen mensen dus hun medewerking weigeren, zoals blijkt uit de situatie rond het ontslag bij Google van Timnit Gebru. Zij was werkzaam in Googles Ethical Artificial en richtte zich als wetenschapper op bias en datamining. Toen zij een paper wilde publiceren over bias in taalmodellen, raakte zij in conflict met haar werkgever, die haar vroeg het paper niet te publiceren of de namen van alle Google-werknemers eruit te verwijderen. Toen zij weigerde, werd zij ontslagen en dat leidde tot verontwaardiging. Duizenden werknemers van het bedrijf, wetenschappers en partijen uit het maatschappelijk middenveld ondertekenden een brief waarin het ontslag werd afgekeurd. Leden van het Amerikaanse congres vroegen Google om uitleg over de zaak.

Technologiebedrijven zijn afhankelijk van getalenteerd personeel, dat daarmee een pressiemiddel in handen heeft om het bedrijfsbeleid te beïnvloeden. Dat geldt zelfs voor toekomstige potentiële werknemers. In 2018 besloten studenten aan de Stanford-universiteit, een prestigieus instituut op het gebied van AI, bijvoorbeeld om niet meer met Google in gesprek te gaan totdat het bedrijf het genoemde Project Maven stopzette. Studenten protesteerden ook tegen de werving op campussen door bedrijven die grenscontroles en politieactiviteiten ondersteunen. Meer dan 1.200 studenten van zeventien campussen tekenden een belofte om niet te gaan werken bij Palantir vanwege de banden die het bedrijf heeft met ICE. En aan Central Michigan University bestreden studenten de oprichting van een universitaire Army AI Task Force.

Engagement in de vorm van uitstappen lijkt in ieder geval in de eerste fase van de maatschappelijke inbedding van een technologie een belangrijke manier voor partijen om hun stem te laten horen. Het aantal toepassingen van AI groeit in rap tempo en werknemers, studenten en platformmedewerkers zitten als het ware dicht bij het vuur. Door op basis van hun kennis van ontwikkelingen actie te voeren vervullen zij een belangrijke signalerende functie voor problematisch gebruik van AI. En daar waar zij succesvol zijn met hun walkout, bijvoorbeeld doordat de rechter die erkent, heeft hun engagement een corrigerende werking.

574 Frenkel, 19 juni 2018.

Een geïnstitutionaliseerder vorm van protest is wanneer vakbonden oproepen tot een staking. Het stakingsrecht is als fundamenteel recht verankerd in het Europees Sociaal Handvest (art. 6, lid 4) en vloeit mede voort uit de vrijheden van vereniging en vergadering. Zo ging enkele jaren geleden een deel van de Deliveroo-bezorgers in staking voor betere arbeidsomstandigheden, waarbij zij werden ondersteund door de binnen de FNV opgerichte 'Riders Union'. Mede onder druk van deze en andere acties kwamen de werkomstandigheden van platformwerkers ook in Nederland op de politieke agenda en wordt er nu gewerkt aan een betere wettelijke bescherming.[575]

Protest: campagne voor verbod

Protest is de derde vorm van verzet die zich tegen een technologie of een bepaald gebruik daarvan keert. Juist in een democratische samenleving komt deze vorm regelmatig tot uiting. Burgers mobiliseren zich hierbij vreedzaam om de autoriteiten op te roepen bijvoorbeeld een verbod in te stellen. Protest is daarbij, in tegenstelling tot de bovenstaande voorbeelden van uitstappen, niet van binnenuit georganiseerd en gericht op het beleid van bedrijven, maar richt zich op de overheid en heeft vaak een bredere maatschappelijke basis. Denk wederom aan de anti-kernenergiebeweging die ook op vreedzame wijze protesteerde om de overheid op te roepen geen kerncentrales te bouwen. Of denk aan allerlei protesten tegen militaire technologie als chemische wapens en clusterbommen. Dergelijke brede burgerbewegingen hebben uiteindelijk vaak geleid tot internationale verdragen om bepaalde wapens te verbieden.

Tekstbox 6.2 – AI bij de Nederlandse politie

Ook in Nederland zijn er vragen gesteld over het gebruik van AI door de politie. Naar aanleiding van de Kamerbrief van de vaste commissie voor Justitie en Veiligheid van 3 december 2019 inzake artificiële intelligentie bij de politie schreef de minister voor Rechtsbescherming onder meer:

> "Slechts in een aantal gevallen ontwikkelt en gebruikt de politie op dit moment AI. Dit aantal zal in de toekomst gaan groeien. Zoals ik in genoemde brief uiteen heb gezet is het gebruik van AI bij de politie geen luxe maar noodzaak om effectief te kunnen blijven optreden. (…) Predictive policing wordt binnen de politie gebruikt als hulpmiddel om een inschatting te maken of er een (verhoogd) risico is op bepaalde vormen van criminaliteit in een bepaald gebied zodat hierop vooraf kan worden geanticipeerd bij de inzet van de gebiedsgebonden politie. Het is geen opsporingsmethode en wordt ook niet gebruikt voor opsporen.

575 Houwerzijl 2018.

> Bij predictive policing is er dus geen sprake van een redelijk vermoeden van een strafbaar feit of een verdenking. Het risico wordt bepaald op basis van historische data, waargenomen trends en patronen waarbij alleen wordt gekeken naar het optreden van een (verhoogd) risico in een bepaald gebied. Het is niet gericht op individuen of groepen.
>
> (…) De politie maakt gebruik van (klassieke) risicotaxatiemodellen om inschattingen te kunnen maken ten aanzien van het potentieel gebruik maken van geweld of het optreden van recidive bij personen met antecedenten zoals vastgelegd in de politiesystemen. Het taxatiemodel geeft alleen een inschatting of er een verhoogd risico is op het gebruik van geweld of het optreden van recidive bij een geselecteerde groep personen."[576]

Ten aanzien van AI is protest een van de meest in het oog springende vormen van engagement. Dat geldt in het bijzonder voor drie toepassingen van AI: gebruik door de politie voor surveillance en voorspellingen, gezichtsherkenning en autonome wapens. Een prominente beweging tegen het gebruik van AI bij politiediensten ontstond een paar jaar geleden in Los Angeles. Een community-groep spande een rechtszaak aan en zorgde ervoor dat de Los Angeles Police Department het *predictive policing*-systeem 'LASER' niet langer gebruikte. Deze groep noemde zichzelf 'the Stop LAPD Spying Coalition'. Zij beargumenteren dat de politie op oneerlijke wijze misdaden voorspelde, door middel van proxy-data, en zo mensen uit de Latino- en zwarte gemeenschap discrimineerde. Studenten van de universiteit van California (UCLA) voegden zich bij deze beweging, en gebruikten onderzoek van de UCLA naar het *predictive policing*-programma PredPol om de beweging te ondersteunen. Uit dat onderzoek bleek dat dit soort instrumenten leidde tot overmatige politie-inzet in gekleurde gemeenschappen.

In St. Louis, Missouri, demonstreerden bewoners ook tegen politietechnologie en in het bijzonder tegen een samenwerking tussen de politie van St. Louis en een bedrijf genaamd Predictive Surveillance Systems. Dit bedrijf zet bewakingsvliegtuigen of drones in om beelden van burgers te verzamelen. De bewoners kwamen in opstand omdat zij stelden dat deze "*suspicionless tracking*" een grote inbreuk zou zijn op hun privacy. In Nederland riep Amnesty ertoe op het politieproject 'Sensing' in een winkelcentrum in Roermond stop te zetten. Slimme camera's werden daar ingezet om mobiel banditisme te bestrijden, maar

576 Kamerstukken II 2019/20, 26 643, nr. 652.

volgens Amnesty was hier sprake van massasurveillance en werd er gediscrimineerd tegen bepaalde groepen op basis van hun nationaliteit.[577]

Een tweede toepassing van AI waartegen veel geprotesteerd wordt, is gezichtsherkenning. Steeds meer camera's zijn toegerust met deze vorm van computer vision en dat maakt het mogelijk concrete individuen heel precies te monitoren. Bezorgde burgers zien hierin een instrument voor totalitaire surveillance. Sommigen streven dan ook naar een verbod op elk gebruik van gezichtsherkenning, anderen benadrukken dat vooral overheden er geen gebruik van zouden mogen maken en weer anderen willen het gebruik aan zeer strenge eisen binden, zoals het niet opslaan van data of het gebruik van het instrument alleen voor bijvoorbeeld vermiste kinderen. Naast de algemene zorgen over surveillance is er ook de zorg dat deze technologie minder goed werkt voor mensen uit minderheidsgroepen en vooral zal bijdragen aan de onderdrukking van die groepen.

Er zijn wereldwijd verschillende maatschappelijke organisaties ontstaan die demonstreren tegen het gebruik van gezichtsherkenning (zie ook tekstbox 6.3). Zo is er Stop Secret Spy Tech and Face Surveillance San Francisco. Zulke protesten in verschillende Amerikaanse steden zijn ook succesvol geweest. In San Francisco en Boston mag de politie deze technologie niet gebruiken en in Portland is zelfs elk gebruik ervan verboden. WHY ID, Electronic Frontier Alliance en Public Voice zijn andere voorbeelden van Amerikaanse bewegingen die tegen het gebruik van gezichtsherkenning protesteren. Ook Europese organisaties als Privacy International (Verenigd Koninkrijk) en Techno Police (Frankrijk) zetten zich hiervoor in. In Nederland roept Bits of Freedom op tot een verbod op gezichtsherkenning en andere biometrische surveillancetechnologie zoals geluidsherkenning.[578]

Tekstbox 6.3 – Een verbod op gezichtsherkenning?

Ook in aanloop naar het concept voor de Europese AI-Verordening is gepleit voor een verbod op gezichtsherkenning. De oproep hiertoe werd gedaan door tientallen civilsociety-organisaties.[579] Daarnaast hebben meer dan 60 Europarlementariërs en ruim 50.000 EU-burgers de oproep ondersteund.[580] De ondertekenaars bepleitten onder meer:

577 Amnesty International 2020b.
578 Bits of Freedom, z.d.
579 Een campagne genaamd 'Reclaim your face'.
580 Reclaim Your Face, 16 april 2021.

1. Een verbod op het zonder onderscheid of arbitraire wijze gebruiken van biometrische identificatie in openbare of openbaar toegankelijke ruimten, hetgeen kan leiden tot massasurveillance;
2. Wettelijke beperkingen of harde grenzen aan gebruik dat fundamentele rechten onder druk zet, zoals AI-toepassingen voor grenscontrole, *predictive policing*, de toegang tot socialezekerheidssystemen en risico-inschattingen in het kader van het strafrecht.

Vanuit Nederland ondertekenden Bits of Freedom, en Het Nederlands Juristen Comité voor de Mensenrechten (NJCM) de oproep. Deze lijkt effect te hebben gehad. In de concept-Verordening AI zijn onder meer *social scoring* en biometrische identificatiesystemen in de openbare ruimte verboden, omdat zij naar het oordeel van de Europese Commissie een 'onacceptabel risico' vormen voor de Europese waarden.

Een derde AI-gerelateerd onderwerp waarbij protestbewegingen mondiaal oproepen tot een verbod, zijn autonome wapens. In 2012 werd de Campaign to Stop Killer Robots opgericht, een coalitie van niet-gouvernementele organisaties die zich inzetten om volledig autonome wapens te verbieden en die pleit voor het behoud van zinvolle menselijke controle over het gebruik van geweld. Meer dan duizend experts op het gebied van AI, waaronder Stephen Hawkins, Elon Musk, Steve Wozniak en Noam Chomsky, tekenden in 2015 een open brief waarin gewaarschuwd wordt voor een AI-wapenwedloop en opgeroepen wordt tot een verbod op autonome wapens. In 2017 werd een vergelijkbare open brief naar de Verenigde Naties gestuurd waarin gepleit wordt voor een verbod op dodelijke autonome wapens. De brief is ondertekend door 166 pioniers in de robotica en directeuren van verschillende technologiebedrijven. In Nederland verscheen eind 2020 een brief aan de Nederlandse regering met eenzelfde boodschap. Meer dan 150 wetenschappers op het gebied van robotica en AI vragen hierin de regering om mee te werken aan een verbod op dodelijke autonome wapens. Ook de Nederlandse organisatie PAX, die zich inzet voor vrede, stelt een verbod voor op de ontwikkeling, de productie en het gebruik van '*killer robots*'. Protest is dus een belangrijke vorm van engagement in een democratische samenleving. Deze vorm is al sterk ontwikkeld ten aanzien van AI en zal prominent blijven.

Kernpunten – Verzet: vechten, uitstappen, protest

- Een nieuwe technologie roept vaak verzet op, zeker wanneer technologische verandering snel gaat en het idee postvat dat de baten van de technologie slechts ten goede komen aan een zeer beperkt deel van de samenleving, terwijl de risico's wijdverspreid zijn. *Verzet* drukt een antagonistische houding tot AI uit en kent verschillende uitdrukkingsvormen: vechten, uitstappen en protest.
- In het verleden hebben groepen zich regelmatig met geweld tegen nieuwe technologie verzet; bij AI zien we deze vorm van *vechten* vooralsnog niet optreden. Deze problematische vorm van engagement behoeft geen versterking – democratische betrokkenheid bij AI verdient de voorkeur.
- Bij *uitstappen* weigeren mensen op verschillende manieren hun medewerking aan AI, bijvoorbeeld door hun werk neer te leggen ('walkouts') en zodoende bedrijven van binnenuit te dwingen van koers te veranderen. De laatste jaren groeit deze vorm van engagement, die typerend is voor de beginfase van AI.
- Een sterk ontwikkelde en voor de democratische samenleving belangrijke vorm van verzet is *protest*. Burgers mobiliseren zich hierbij vreedzaam om de autoriteiten op te roepen ergens een verbod op in te stellen. Het is momenteel een van de meest in het oog springende vormen van engagement, die zich met name richt tegen drie toepassingen van AI: gebruik door de politie voor surveillance en voorspellingen, gezichtsherkenning en autonome wapens.

6.2 Monitoren

Het tweede cluster van vormen van engagement dat we onderscheiden, is 'monitoren'. Dit cluster bevindt zich in het midden van het spectrum van enerzijds de antagonistische vormen van engagement en anderzijds de symbiotische vormen, die we in de volgende paragraaf zullen bespreken. Als het gaat om monitoren, onderscheiden we twee vormen van engagement: controleren en agenderen. In beide gevallen gaat het om een houding waarbij het handelen van andere partijen – zowel publiek als privaat – kritisch wordt gevolgd en waar nodig ook gecorrigeerd en bijgestuurd met alternatieve voorstellen. Deze houding sluit aan bij de historische wending die John Keane signaleerde in *The Life and Death of Democracy*, en die erop neer komt dat er na 1945 vele honderden nieuwe typen instituties zijn ontstaan die het handelen van invloedrijke partijen volgen en aan nauwkeurig onderzoek onderwerpen.[581] Keane, die deze ontwikkeling typeert als *'monitory democracy'*, wijst daarbij

581 Keane 2009.

onder meer op het gebruik van enquêtes, onlinepetities en focusgroepen maar ook op zelfbenoemde '*watchdogs*' en ngo's begaan met zwakkere of ondervertegenwoordigde groepen in de samenleving.[582] Hierop voortbouwend verstaan we onder 'controleren' de organisatie van belanghebbenden om misstanden in het gebruik van de nieuwe technologie aan te kaarten. De vijfde en wat neutralere vorm van engagement, duiden wij als agenderen. Het gaat hierbij om partijen uit het maatschappelijk middenveld die zowel positieve als negatieve kanten aan de technologie onderscheiden, maar zich vooral inzetten om het thema maatschappelijk meer aandacht te geven.

Controleren: misstanden aankaarten

De doelstelling van controleren is anders dan de hiervoor besproken vormen van engagement. Het gaat bij controle niet zozeer om het voorkomen van specifieke toepassingen van AI – een verbod – maar om een correctie van die toepassingen of de condities waaronder die toepassingen plaatsvinden. Bijvoorbeeld door specifieke partijen te informeren of door publieke campagnes, maar ook via rechtszaken en meldingen bij toezichthouders om zo misstanden te adresseren. Rechten – mensenrechten voorop – zijn daarbij in de praktijk een cruciaal referentiepunt. Een belangrijke functie van controleren is dat partijen uit het maatschappelijk middenveld AI-toepassingen toetsen op hun wettelijk toegestane karakter.

Er is in dit vroege stadium van maatschappelijke inbedding onzekerheid over de effecten van AI. Rechtszaken vervullen een belangrijke rol om met die onzekerheid om te gaan. Via die weg worden misstanden geïdentificeerd en jurisprudentie ontwikkeld. Daarmee kunnen vanuit rechtsherstel de direct benadeelde groepen beschermd worden, en wordt bovendien de rechtsontwikkeling gediend. Jurisprudentie heeft daarmee ook een signalerende en sturende functie. Zij biedt zicht op wat er in het veld gebeurt, levert duidelijkheid op om daarop in te spelen en kan tevens bijdragen aan een kader voor verdere toepassingen en mogelijk ook toekomstige wetgeving. De geschiedenis leert dat rechtszaken tegen misbruik van spoorwegen en telegrafen deze rol expliciet hebben gespeeld.[583]

Hiernaast zijn er situaties waarin AI-toepassingen verplicht aan een 'toets' onderworpen dienen te worden, doordat belanghebbenden gehoord moeten worden over de ingebruikname van nieuwe technologische mogelijkheden. Dat is bijvoorbeeld het geval via bij het adviesrecht (over bijvoorbeeld investeringen), het instemmingsrecht en het informatierecht dat ondernemingsraden hebben. Het belang van deze rol blijkt nog eens uit de handreiking die de

582 Keane 2011.
583 Pasquale 2015: 90-1.

Sociaal-Economische Raad (SER) uitbracht om ondernemingsraden te informeren over hun rol als het gaat om technologische ontwikkelingen.[584]

Controlerende activiteiten kunnen ondernomen worden door belangengroepen, experts of door de media. Deze vorm van engagement is veelal wat minder publiek zichtbaar dan protesten, maar is eveneens een belangrijke vorm die bovendien steeds groter wordt. Dat laatste komt met name door de opkomst van relatief nieuwe organisaties die zijn ontstaan rondom het gebruik van AI in de samenleving en breder rondom de opkomst van digitalisering. In Nederland speelt bovendien dat redelijk recent de mogelijkheid tot collectieve actie bij gerechtelijke procedures wettelijk is verruimd.[585] Daardoor wordt deze vorm van engagement gemakkelijker. Een regelmatig genoemd punt van zorg is desalniettemin dat AI-gebruik op gespannen voet staat met de huidige juridische bescherming van slachtoffers, omdat deze op individuen is gebaseerd, terwijl AI-toepassingen individuen in groepsprofielen onderbrengen.[586]

We bespreken hier eerst een aantal organisaties op het internationale en nationale niveau die vooral via kennisdeling bijdragen aan de controle op het gebruik van AI. Daarna staan wij stil bij een aantal prominente rechtszaken.

Internationaal zijn er verschillende organisaties die kennis verspreiden over het gebruik van AI en daarmee een controlerende rol vervullen. In 2017 richtten Kate Crawford en Meredith Whittaker in New York het AI Now instituut op. Door middel van rapporten en analyses houdt dit instituut zich bezig met de effecten van AI op vier terreinen: rechten en vrijheden, werk en automatisering, vooroordelen en inclusie, en veiligheid en infrastructuur. De organisatie heeft misstanden aangekaart rond de slechte werkomstandigheden van mensen in de technologiesector (bijvoorbeeld in de magazijnen van Amazon) maar ook rond de ecologische footprint van AI-systemen, waar weinig aandacht voor is. In een jaarlijks rapport beschrijft AI Now de huidige stand van zaken met betrekking tot het gebruik van AI en doet het instituut aanbevelingen voor de verdere ontwikkeling van AI in de samenleving.

Een andere organisatie is het in 2016 gelanceerd Google Transparency Project. Dat was erop gericht om met onderzoek en analyse licht te werpen op de manier waarop het bedrijf Google de overheid en het beleid beïnvloedt. Inmiddels is de organisatie omgedoopt tot Tech Transparency Project en richt het zich breder op

584 SER 2016. Deze rol is, zoals ook blijkt uit de SER-handreiking, natuurlijk veel breder dan controlerend, maar de laatstgenoemde is daar wel nadrukkelijk een onderdeel van.
585 Van Boom en Weder 2017. Het betreft de Wet afwikkeling massaschade in collectieve actie.
586 Kosta 2020; Van der Sloot en Van Schendel 2020; Taylor et al. 2016.

de technologiesector. Deze non-profitwaakhond streeft daarbij verantwoording van bedrijven na middels onderzoek, procesvoering en het publiek maken van wangedrag.

Een belangrijke partij in Europa is het Duitse AlgorithmWatch. Deze non-profitorganisatie is gericht op algoritmische besluitvormingsprocessen die maatschappelijke effecten hebben, zoals algoritmes om gedrag te voorspellen of te sturen, of algoritmes die automatisch besluiten kunnen nemen. Dat doet de organisatie door te analyseren hoe algoritmische besluitvorming het menselijke gedrag beïnvloedt, door aan het grote publiek uitleg te geven over hoe die besluitvorming werkt, door experts bijeen te brengen en door ideeën en strategieën te ontwikkelen om AI op een goede manier in de samenleving te leren gebruiken. Met het jaarlijkse *Automating Society Report* brengt AlgorithmWatch het gebruik van automatische besluitvorming in Europa in kaart, waaronder in Nederland. Ook brengt het insituut deelstudies uit naar bijvoorbeeld het gebruik van algoritmische besluitvorming als reactie op de coronacrisis. De organisatie brengt ethische dilemma's in kaart en draagt ook zelf voorstellen aan voor een verantwoorder gebruik van algoritmes.

In Nederland zijn er ook verschillende organisaties die zich expliciet bezighouden met de maatschappelijke effecten van AI en aanpalende technologieën. Dat geldt bijvoorbeeld voor de door Marleen Stikker geleide Waag Society en voor Privacy First. Waag Society opereert als onderzoeksinstelling op het kruispunt van kunst, wetenschap en technologie. De organisatie voert onderzoek en ontwikkeling uit rond technologische en maatschappelijke vraagstukken. Daarnaast organiseert Waag publieksprogramma's met workshops, tentoonstellingen en debat. De Waag doet studie naar kunstmatige intelligentie en zet zich in om de samenleving bewust te maken van deze technologie. Privacy First zet zich in voor behoud en bevordering van het recht op privacy. De activiteiten van Privacy First variëren van politieke lobby, juridische acties en rechtszaken, informatieverstrekking en campagnes voor het grote publiek. Privacy First streeft ernaar continu de discussie te voeren in de Nederlandse samenleving.

Dergelijke organisaties dragen bij aan de controle van nieuwe technologieën door kennis te leveren. Zo illustreerde de digitale burgerrechtenorganisatie Bits of Freedom dat het wel degelijk mogelijk was om vanuit een ander land op Facebook te adverteren, bijvoorbeeld tijdens verkiezingen in dat land. Zo kon Bits of Freedom met gemak vanuit Duitsland in Nederland politieke memes, ofwel ideeën, uploaden die bepaalde politieke boodschappen van partijen uitdragen.[587] Dit ondanks het feit dat Facebook de Tweede Kamer anders vertelde.

587 Austin, 20 mei 2019

Partijen kunnen echter ook een misstand adresseren door naar de rechter te stappen. Zo daagde een coalitie van betrokkenen uit het Nederlandse maatschappelijk middenveld de Nederlandse staat voor de rechter om een verbod te eisen op de inzet van het Systeem Risico Indicatie (SyRI). Dit systeem werd in samenwerking met het ministerie van Sociale Zaken en Werkgelegenheid (SZW) door verschillende gemeenten ingezet om onder meer gevallen van bijstandsfraude op te sporen. De coalitie bestond uit de maatschappelijke organisaties Platform Bescherming Burgerrechten, het Nederlands Juristen Comité voor de Mensenrechten, vakbond FNV, Privacy First, Stichting KDVP, de Landelijke Cliëntenraad en de schrijvers Tommy Wieringa en Maxim Februari.

De coalitie stelde dat er bij SyRI sprake was van onrechtmatige *automated decision making*. Ook voerden de eisers aan dat SyRI vooral werd ingezet in wijken die al als probleemwijk bekend staan, waardoor het systeem een discriminerend en stigmatiserend effect heeft. De rechtbank stond in haar oordeel eerst stil bij het karakter van SyRI, dat volgens haar aansloot bij vormen van AI als 'deep learning' en zelflerende systemen. Gelet op de omstandigheid dat SyRI gebruik maakt van risicoprofielen, kan de inzet ervan volgens de rechtbank (onbedoeld) leiden tot verbanden door bias, bijvoorbeeld op basis van een lagere sociaaleconomische status of een immigratieachtergrond. SyRI raakt arme burgers daarmee op disproportionele wijze. Het belang om uitkeringsfraude op te sporen verhoudt zich volgens het oordeel van de rechter niet tot de inbreuk die de staat met het systeem maakt op de privacy van zijn burgers. Als gevolg van de uitspraak werd het stopgezet.

Tekstbox 6.4 – Systeem Risico Indicatie

Het Systeem Risico Indicatie (SyRI) is een technische toepassing die berekent hoe groot de kans is dat een bepaald individu mogelijk fraudeert met sociale uitkeringen. Daarvoor koppelt SyRI 17 soorten gegevens aan elkaar. SyRI vergelijkt volgens het kabinet bestanden met bestaande, feitelijke gegevens van onder meer UWV, SVB, gemeenten/colleges van B&W, Belastingdienst, en Inspectie SZW. Vervolgens beoordeelt het systeem of er discrepanties zijn tussen de verschillende gegevens. Als uit die onderlinge vergelijking en toetsing aan het risicomodel een discrepantie volgt, moet deze eerst door een of meer van de genoemde partijen worden onderzocht. Pas daarna mag er een beslissing worden genomen die voor de betrokkene rechtsgevolgen kan hebben.

Inmiddels kijken de betrokken partijen ook naar het voorstel voor de Wet Gegevensuitwisseling in samenwerkingsverbanden (WGS). Deze wet moet volgens het kabinet de gegevensverwerking door samenwerkingsverbanden van publieke en private organisaties van een juridische basis voorzien. Met het wetsvoorstel wilde het kabinet-Rutte III niet alleen de aanpak van fraude versterken, maar ook de mogelijkheden vergroten om in samenwerkingsverbanden gegevens te delen rondom de aanpak van ondermijning. De vrees bestaat dat de wet nog omvangrijkere gegevensanalyses mogelijk maakt, nu de overheid daartoe ook met private partijen gaat samenwerken. Tegenstanders noemen de wet daarom ook wel 'Super SyRI'. Het gaat onder de WGS volgens hen niet alleen om feitelijke gegevens die bedrijven en overheden met elkaar delen, maar ook om signalen, vermoedens en volledige zwarte lijsten die worden uitgewisseld en met elkaar worden verknoopt. Daarbij zou het de bedoeling zijn dat deze partijen op basis van deze schaduwadministraties 'interventies' met elkaar afstemmen waarin ze handhavend optreden tegen burgers die ze in het vizier krijgen. Een ander pijnpunt is volgens hen dat sommige kwesties worden bepaald zonder dat de Kamer daarmee instemt. Daarbij kan het bijvoorbeeld gaan om de hoeveelheid en soorten gegevens, de partijen die erbij kunnen en de manier waarop ze worden geanalyseerd en verder worden gebruikt.

Het is niet alleen de overheid die zich in rechtszaken moet verantwoorden over het AI-gebruik. Ook private bedrijven zoals Uber zijn vanwege allerlei misstanden aangeklaagd. Zo heeft FNV Uber gedagvaard omdat het bedrijf chauffeurs niet volgens de cao voor taxichauffeurs betaalt. In een andere zaak stelde de rechter Uber in het gelijk door vast te stellen dat het bedrijf rechtmatig gebruik maakte van algoritmes om te bepalen welke chauffeurs wel of niet werden ontslagen.[588]

Overigens is het lang niet altijd eenvoudig voor partijen in het maatschappelijk middenveld om een zaak voor de rechter te brengen. Ze moeten de middelen en de kennis hebben om misstanden aan te kaarten en tot actie over te gaan. Juist de traditionelere belangenorganisaties, veelal ontstaan in het analoge tijdperk, ontbreekt het bovendien regelmatig nog aan bewustzijn over de manier waarop AI hun werkveld verandert.[589] Zij doorzien bijvoorbeeld nog onvoldoende hoe AI de door hen vertegenwoordigde groepen kan marginaliseren of hun belangen onder druk kan zetten. Neem bijvoorbeeld de consument, die in economische transacties met een bedrijf als een zwakke partij wordt beschouwd en daarom in Europa juridische bescherming krijgt. Het gebruik van AI-systemen kan een impact hebben op de autonomie van consumenten, omdat niet de consument

588 Zie: Rechtbank Amsterdam, 11 maart 2021.
589 Steijns 2021.

zelf maar een algoritme op basis van een geïdentificeerde behoefte op zoek gaat naar een optimale aankoop.[590] Bij gebrek aan informatie over de achterliggende werking van AI-systemen kunnen bedrijven consumenten bovendien overtuigen om aankopen te doen die niet in hun belang zijn, bijvoorbeeld omdat ze duurder zijn. Hierdoor dreigt de grens tussen gepersonaliseerde aanbiedingen en manipulatie te vervagen.[591] Behalve aan de wetgever is het bijvoorbeeld ook aan consumentenorganisaties om op een dergelijke ontwikkeling zicht te krijgen en daartegen in het geweer te komen, mocht dat nodig zijn.

Bijzonder aan de rechtszaak tegen SyRI was dat deze gedreven werd door een coalitie van verschillende soorten partijen, waaronder enkele traditionele belangenorganisaties en experts op het gebied van het recht en digitale technologie. Dergelijke coalities zijn erg behulpzaam om de functie van controle door het maatschappelijk middenveld goed invulling te geven. Organisaties die minder goed bekend zijn met AI en de problemen die met digitalisering samenhangen, kunnen gebruik maken van de expertise van organisaties die deze kennis wel hebben. Naast digitale rechtenorganisaties zijn het in toenemende mate ook mensenrechtenorganisaties die deze kennis en expertise in bezit hebben en verder ontwikkelen.[592] Zowel de mensenrechtenorganisaties als digitale rechtenorganisaties maken bovendien deel uit van grotere internationale netwerken, waar AI al langer op de agenda staat.

Het samenkomen van dit soort organisaties voor het voeren van een rechtszaak wordt gestimuleerd door een type procedure dat ook wel bekend staat als strategisch procederende belangenorganisaties (of *public interest litigation)*.[593] Het Nederlandse rechtssysteem is hiervoor echter niet heel ontvankelijk. Organisaties moeten bijvoorbeeld kunnen aantonen dat regels of beleid rechtstreeks de door hun vertegenwoordigde collectieve belangen of het algemeen belang raken. Deze vorm van procederen wordt dan ook nog niet systematisch ingezet. Tegelijkertijd bepleiten juristen verdergaande innovatie in Nederland, specifiek ook met het oog op de voortschrijdende digitalisering. Zo wijzen ze onder meer op de Duitse mogelijkheid voor concurrenten om elkaar via de rechter op de naleving van regels aan te spreken.[594]

Agenderen: informatie over het belang van AI

De volgende vorm van engagement met AI die we onderscheiden, bevindt zich al iets meer aan de rechterkant van het spectrum van antagonisme naar

590 Fierens et al. 2021: 974-975.
591 Fierens et al. 2021: 969.
592 Steijns 2021.
593 Braun en Stolk 2020.
594 Moerel en Prins 2016: 116; en recent Barkhuysen 2021.

symbiose. Het gaat hier om partijen die zich inzetten om meer aandacht voor AI te genereren omdat die aandacht als zodanig belangrijk is. De aandacht kan zijn gericht op positieve en op negatieve kanten van AI. Verschillende partijen uit het maatschappelijk middenveld dragen bij aan het agenderen van AI. Er zijn specifieke organisaties die aandacht vragen voor nieuwe technologie, maar ook opiniemakers en kunstenaars spelen hier een rol. De podia die zij daarbij gebruiken, verschillen sterk van elkaar. We bespreken hier naast kunstuitingen en rapporten van denktanks ook hoe partijen uit het maatschappelijk middenveld betrokken zijn bij de totstandkoming van AI-beleid en AI-wetgeving, en aandacht vragen voor de belangen die zij vertegenwoordigen.

Veel van de bovengenoemde internationale organisaties als AI Now en AlgorithmWatch controleren niet alleen AI, maar informeren middels rapporten en evenementen ook over AI. Kunstenaar Trevor Paglen maakte met het ImageNet Roulette een app waarmee mensen foto's van hun gezichten konden uploaden om te zien hoe zij geclassificeerd worden door de invloedrijke databank ImageNet.[595] Een Nederlandse organisatie die hier bij uitstek aan bijdraagt, is de al eerder genoemde onderzoeksinstelling Waag. Met bronnen in de hackersbeweging en de uitrol van het internet in Nederland richt deze organisatie zich op het realiseren van een open, eerlijk en inclusief gebruik van digitale technologie. Ze plaatst publieke waarden en belangen tegenover de invloed van commerciële logica.

Maxim Februari is een NRC-columnist die regelmatig aandacht vraagt voor het gebruik van technologieën als AI. Hij was ook onderdeel van de groep die de rechtszaak tegen SyRI aanspande en draagt dus ook bij aan het controleren van AI(-gebruik) en het aankaarten van misstanden. Door in zijn columns regelmatig te schrijven over onderwerpen als dataverzameling en algoritmes zorgt hij ervoor dat vraagstukken rondom AI ook breder geagendeerd worden. Dat geldt ook voor het schrijven van fictie. Leidmotief van zijn dystopische roman *Klont* is 'kennen of gekend worden', wat slaat op het idee dat wij de data niet gebruiken maar via data vooral zelf gebruikt worden en achter de 'data' aanlopen. Dat drukt hij uit in de verhouding tussen de twee personages: een vlotte 'spreker' uit het tech-optimistische lezingencircuit, die zelf niet helemaal doorheeft waar hij nu eigenlijk voor pleit, en een wat al te bescheiden/angstige ambtenaar die de tech-optimist voor de minister van justitie in de gaten moet houden.[596] Ook andere verschillende Nederlandse kunstprojecten zijn erop gericht om bewustwording over AI te vergroten. In de tekstbox 6.5 bespreken we er twee.

595 Crawford 2021: 141.
596 Pruis, 11 oktober 2017.

Tekstbox 6.5 – Agenderen via kunst

We Are Data
Dit kunstenaarscollectief draagt bij aan een dieper bewustzijn van wat er aan persoonlijke gegevens kan worden geregistreerd. Wanneer je dit aan den lijve ervaart, kun je de impact van (nieuwe) technologieën beter begrijpen, zo is het idee. We Are Data ontwikkelde hiertoe de MIRROR ROOM, een afgesloten ruimte waarin één bezoeker per keer een persoonlijke en indrukwekkende ervaring ondergaat. Het is ook een slimme ruimte waarin bezoekers onbewust bekeken en gemeten worden. De bezoeker ervaart hoe zij tot data wordt verwerkt en krijgt de keuze welke persoonlijke gegevens zijn of haar eigendom blijven. Op die manier wordt de bezoeker letterlijk een spiegel voorgehouden.

Wouter Moraal
Moraal wil met zijn werk mensen voorlichten over hoe *deep learning*-algoritmes werken en ze waarschuwen voor mogelijke gevolgen van misbruik daarvan. Hij ontwikkelde hiertoe het Artificial Impact bordspel, een bordspel in de vorm van een deep learning-algoritme.

Tijdens het spel moeten de spelers eerst het algoritme trainen. Aan het einde van het spel worden ze zelf beoordeeld door hun eigen creatie, een autodidactisch algoritme voor risicovoorspelling. De beoordelingen in het spel zijn gebaseerd op situaties waarin AI onzorgvuldig werd gebruikt en tot misstanden leidde. Dit project is daarmee vergelijkbaar met het bordspel Monopolie dat in 1903 door Lizzie Magie werd ontwikkeld om mensen bewust te maken van de kwalijke gevolgen van grootgrondbezit en kapitalistische uitbuiting.

Een laatste voorbeeld van de hier besproken vorm van engagement is DenkWerk. Dat is een onafhankelijke denktank met leden uit verschillende domeinen van de overheid tot het bedrijfsleven en de wetenschap. Zij richten zich op het agenderen van brede maatschappelijke vraagstukken en het aandragen van oplossingen in periodieke rapporten. Die rapporten gaan over uiteenlopende onderwerpen, maar specifiek op het onderwerp van AI is DenkWerk van belang geweest. Zo zwengelde DenkWerk in een in juli 2018 verschenen rapport over AI in een vroeg stadium in Nederland het debat aan over het potentieel van de technologie en wat ervoor nodig is om dat te realiseren.[597] Daarmee

597 DenkWerk 2018.

droeg DenkWerk bij aan het momentum om te komen tot een Nederlandse AI-strategie. Ook het latere rapport *De Onlinewereld.nl* agendeert AI.

Deze vorm van engagement is belangrijk, maar in Nederland nog maar beperkt ontwikkeld en bereikt vooral een al geïnteresseerd publiek. Belangrijk voor de toekomst is om met het agenderen van AI juist bredere lagen van de bevolking te bereiken.

Het agenderen van de verschillende kanten van AI is met name ook belangrijk bij politieke besluitvormingsprocessen. In een representatieve democratie hebben politieke vertegenwoordigers het laatste woord bij de totstandkoming van politieke besluiten. Daarbij moeten ze wel verantwoording afleggen aan kiezers en de bredere samenleving. In de praktijk zijn daartoe verschillende processen ingericht, met een meer en minder vrijblijvend karakter. Partijen uit het maatschappelijk middenveld kunnen zelfstandig hun standpunten naar voren brengen aan Kamerleden of departementen die aan beleids- en/of wetsvoorstellen werken. Die voorstellen zijn er ook in toenemende mate als het gaat om AI en AI-gebruik in verschillende sectoren. We bespreken hier de betrokkenheid van partijen uit het maatschappelijk middenveld bij de concept-Verordening AI van de Europese Commissie en het Strategisch Actieplan voor AI (SAPAI) van de Nederlandse regering.

Op 21 april 2021 publiceerde de Europese Commissie de concept-Verordening AI.[598] Bij de totstandkoming daarvan speelden op verschillende momenten ook partijen uit het maatschappelijk middenveld een rol. Allereerst gebeurde dat door deelname aan de AI HLEG die in 2018 werd opgericht. In dit verband adviseerden 52 experts de Europese Commissie over de implementatie van de AI-strategie van de Europese Commissie, die op 7 december 2018 verscheen.[599] Vier van de leden van de AI HLE waren vertegenwoordigers van het maatschappelijk middenveld, naast 18 academici en 37 vertegenwoordigers van het bedrijfsleven. De AI HLE presenteerde op 17 juli 2020 een definitieve beoordelingslijst voor betrouwbare AI.[600] De kritiek vanuit het maatschappelijk middenveld was al tijdens de beraadslagingen van de AI HLE dat de industrie op deze lijst een te grote stempel had weten te drukken en met name een aantal voorstellen om sommige vormen van AI te verbieden had tegengehouden.[601] Een ander relevant kritiekpunt was dat met name de stem ontbrak van partijen met praktijkkennis en van binding met die groepen in de samenleving die met

598 Europese Commissie 2021b.
599 Europese Commissie, 7 december 2018.
600 High-Level Expert Group on AI 2020.
601 Zie in het bijzonder Access Now 2020: 16-18.

AI-systemen te maken krijgen. Volgens Michael Veale, een Britse onderzoeker op het terrein van digitale rechten, zijn het juist deze zogenoemde 'low level experts', die met ethische afwegingen te maken zullen krijgen op het moment dat AI-toepassingen worden geïmplementeerd.[602] Volgens Veale is de noodzaak voor deze praktijkdeskundigen vele malen groter dan voor de door de AI HLEG bepleite hoogleraarsposities toegepaste ethiek.

Het maatschappelijk middenveld was eveneens betrokken via de AI-alliantie, die fungeert als platform voor ongeveer 4.000 stakeholders. Het aanvankelijke doel van het forum was om feedback te verstrekken aan de AI HLEG. In de loop van de tijd is de AI-alliantie echter een referentiepunt geworden in door belanghebbenden aangestuurde discussies over AI-beleid.

Tot slot namen verschillende partijen afkomstig uit het maatschappelijk middenveld deel aan de publieke consultatie voorafgaande aan de publicatie van het Witboek AI, dat op 19 februari 2020 verscheen. 84 procent van de bijdragen was afkomstig uit EU-lidstaten, de overige uit de rest van de wereld. 13 procent van de bijdragen kwam van civil society-actoren.[603] Een groot aantal van hen vond dat de Commissie veel verder moest gaan in de bescherming van mensenrechten, vooral in relatie tot het gebruik van gezichtsherkenning. in tekstbox 6.3 (pagina 270-271) genoemde gezamenlijke oproep van enkele tientallen civil society-organisaties aan de Europese Commissie om bepaalde vormen van AI en AI-gebruik te verbieden, was daarvan een uitdrukking.[604]

De Europese Commissie consulteerde maatschappelijke partijen, waaronder partijen uit het maatschappelijk middenveld, om hen de gelegenheid te geven hun visie uiteen te zetten over het in 2020 gepubliceerde Witboek over artificiële intelligentie. Het agenderen van vraagstukken die van belang zijn voor specifieke groepen in de samenleving is dan ook, zoals we ook in de inleiding op dit hoofdstuk al betoogden, niet uitsluitend de verantwoordelijkheid van partijen in het maatschappelijk middenveld. Bij uitstek de overheid heeft als taak om zo goed als mogelijk recht te doen aan een veelheid van belangen. Zij zal zich derhalve een beeld moeten vormen van de verschillende standpunten die leven over de inbedding van AI in de samenleving en de mogelijke implicaties van de introductie van AI voor verschillende groepen belanghebbenden. Deze taak vertaalt zich bij politieke besluiten onder meer in de formele eisen om belanghebbenden te consulteren of inspraak toe te staan in de politieke besluitvorming.

602 Veale 2020.
603 Europese Commissie, z.d. (b).
604 Reinhold, 22 april 2021.

In aanloop naar het Strategisch Actieplan voor AI (SAPAI) hadden in Nederland belanghebbenden meer betrokken kunnen worden.[605]

Het SAPAI kwam mede tot stand na aandringen van de Nederlandse werkgeversorganisatie VNO/NCW, vijf grote Nederlandse bedrijven en verschillende hoogleraren, ondersteund door het ECP, het platform voor de informatiesamenleving. Bij de door deze partijen ingestelde AI Task Force (later de AI-Coalitie) waren geen andere maatschappelijke organisaties betrokken. Voorafgaande aan de totstandkoming van het SAPAI werden daarentegen diverse belangenorganisaties geraadpleegd over de belangen van patiënten en consumenten, huiseigenaren en privacyrechten. Ook werden verscheidene conferenties georganiseerd, om belanghebbenden in te lichten en hun opvattingen te horen. Een dergelijk proces achtte de Europese Commissie wenselijk, maar het had desalniettemin een bredere en gestructureerdere invulling kunnen krijgen. Zo had het bestaande internetconsultatiemechanisme een manier kunnen zijn om groepen te bereiken die niet op de radar van de overheid stonden. Enerzijds is het begrijpelijk dat, gezien het vroege stadium waarin AI als systeemtechnologie verkeert, voor de overheid niet onmiddellijk duidelijk is welke groepen betrokken moeten worden bij plannen om de impact van de technologie in banen te leiden. Anderzijds zijn het per definitie juist de eerdergenoemde mensen met praktijkkennis die – wellicht met enige assistentie – in de positie zijn om de uitdagingen te benoemen die zich zullen voordoen bij de maatschappelijke inbedding van AI, zeker omdat inmiddels is bewezen dat juist zwakkere of kwetsbare groepen de negatieve effecten van AI-systemen ondervinden. Het is daarom aan te raden dat de overheid actief en structureel een breed spectrum aan groepen uit het maatschappelijk middenveld benadert.[606]

Kernpunten – Monitoren: controleren, agenderen

– Bij *monitoren* gaat het om een houding waarbij het handelen van private en publieke partijen kritisch wordt gevolgd en waar nodig ook wordt gecorrigeerd en bijgestuurd met alternatieve voorstellen. We onderscheiden twee vormen van engagement – controleren en agenderen – die het midden houden tussen enerzijds antagonistische vormen van engagement en anderzijds meer symbiotische vormen ervan.

605 European Center for Not-for-Profit Law 2020.

606 Vergelijk European Center for Not-for-Profit Law 2020: 7. De regering is niet verplicht om wetten, algemene maatregelen van bestuur of ministeriële regelingen voor te leggen aan belanghebbenden. Wel moet daarover een afweging plaatsvinden. In 2007 sprak de regering de ambitie uit om de dialoog en het overleg met de samenleving te versterken, onder andere door internetconsultatie.

- Bij *controleren* gaat het om een correctie van AI-toepassingen of de condities waaronder die toepassingen plaatsvinden. Dat kan zijn door specifieke partijen te informeren of door publieke campagnes, maar ook via rechtszaken en meldingen bij toezichthouders om misstanden te adresseren. Rechten – mensenrechten voorop – zijn daarbij in de praktijk een cruciaal referentiepunt. Zo ontstaat al in een vroeg stadium zicht ontstaat op de effecten van AI.
- Bij *agenderen* zetten partijen uit het maatschappelijk middenveld, maar ook opiniemakers en kunstenaars, zich in voor meer aandacht voor bepaalde aspecten van AI en AI-gebruik. Deze vorm van engagement is belangrijk, maar in Nederland nog maar beperkt ontwikkeld en bereikt vooral een al geïnteresseerd publiek.
- Agenderen is ook belangrijk bij politieke besluitvormingsprocessen, en gebeurt nationaal en internationaal. Het is essentieel dat de overheid hiertoe een breed spectrum aan groepen uit het maatschappelijk middenveld benadert. Juist zwakkere of kwetsbare groepen ondervinden namelijk vaak de negatieve effecten van AI-systemen.

6.3 Samenwerking

Samenwerking is het derde en laatste cluster van vormen van engagement dat we onderscheiden. Het omvat om te beginnen de inzet om de technologie te verbeteren. Denk hierbij aan partijen uit het maatschappelijk middenveld die principes van goed gebruik opstellen of betrokken zijn bij standaardisatieprocessen. Daarnaast omvat samenwerking het toe-eigenen van de nieuwe technologie. Het gaat hierbij om partijen die de technologie in hun bestaande activiteiten opnemen en gebruiken om hun eigen doelen en waarden te realiseren. Voor samenwerking zijn verschillende redenen te noemen, waarvan er enkele samenhangen met het specifieke karakter van AI.

Verbeteren: kennis over goed gebruik

Op het spectrum van engagementsvormen zit de verbetering van AI aan de kant van symbiose. Het gaat hier om mensen die werkzaam zijn op het gebied van AI of een aanverwante of andere relevante expertise hebben. Zij werken aan AI vanuit de overtuiging dat de technologie de samenleving zal verrijken. Zij mobiliseren zich om hun expertise over het onderwerp in te zetten en zo AI en het gebruik ervan te verbeteren. Dat kan zijn in de vorm van het opstellen van principes, het schrijven van open brieven, het ontwikkelen van instrumenten voor goed AI-gebruik (toolkits) of andersoortige publicaties. Veel van de initiatieven van deze vorm van engagement zijn internationaal van karakter. Als het gaat om het opstellen van principes of normen, zijn ook de EU, de VN en verschillende standaardisatieorganisaties actief, die regulerende macht hebben. Hier kijken wij echter uitsluitend naar de bottom-upinitiatieven vanuit

het maatschappelijk middenveld door bijvoorbeeld professionele organisaties, kennisinstellingen of nieuwe non-profitorganisaties.

In 2015 werd een AI-securityconferentie in Puerto Rico gehouden. In een open brief wezen deelnemers vooral op het belang om onderzoek naar AI te verbreden vanuit de gedachte dat AI 'uit het lab' komt. Volgens de deelnemers moeten ethici, filosofen, economen, rechtsgeleerden en cybersecurityonderzoekers meer geëngageerd worden bij de (interdisciplinaire) onderzoeksagenda.[607]

In 2017 organiseerde het Future of Life-instituut de *Asilomar Conference on Beneficial AI*. Honderd mensen, waaronder AI-wetenschappers, economen, filosofen en juristen maar ook politici kwamen daarbij samen om 23 principes voor heilzame AI te ontwikkelen. Die principes zijn geordend naar vragen voor onderzoek naar AI, ethiek en waarden en lange-termijnvraagstukken.[608] Verschillende prominente onderzoekers waren aanwezig op de conferentie en de lijst met principes is onder andere ondertekend door Elon Musk, Nick Bostrom, Demis Hassabis, Yann LeCun, Yoshua Bengio, Stuart Russell en uit Nederland door onder andere Irakli Beridze, Kees Verhoeven en Catelijne Muller. Muller is directeur van ALLAI, een Nederlandse organisatie die zich via onderzoek en samenwerkingsprojecten inspant voor het ontwikkelen van verantwoorde AI (zie tekstbox 6.6).

Tekstbox 6.6 – ALLAI

Tijdens de World Summit AI in 2018 werd de organisatie Alliance for Artificial Intelligence Netherlands (ALLAI) gelanceerd.[609] Daarmee werd Nederland het eerste land in Europa dat beschikt over een onafhankelijke organisatie die zich volledig toelegt op het verantwoordelijk gebruik van AI. ALLAI richt zich onder meer op het ontwikkelen van ethische randvoorwaarden voor AI door middel van projecten, onderzoeken, beleidsadviezen en educatie. Met verantwoordelijke AI als uitgangspunt ambieert ALLAI zowel nationaal als internationaal een omgeving te creëren waarin de vruchten van AI worden geplukt terwijl publieke waarden zoals veiligheid, autonomie en inclusie worden gewaarborgd. Oprichtsters Catelijne Muller, Virginia Dignum en Aimee van Wynsberghe (voorheen leden van de EU High-Level Expert Group on AI) stimuleren daarvoor de samenwerking tussen verschillende belanghebbende partijen op het gebied van AI en betrekken beleidsmakers,

607 Future of Life Institute z.d. (b).
608 Future of Life Institute z.d. (c).
609 Alliance for Artificial Intelligence z.d.

wetenschappers, ondernemers, juristen en consumenten bij hun projecten. Ook in het kader van de coronacrisis verkent de organisatie de mogelijkheden om AI op verantwoorde wijze in te zetten. Hiervoor werken zij samen met beleidsmakers en onderzoekers.

Vanuit de universiteit van Montreal is ook een set ethische principes voor verantwoordelijke AI ontwikkeld. Vijfhonderd academici, burgers en belanghebbenden werden gemobiliseerd om schriftelijk en via bijeenkomsten te reageren op een conceptvoorstel van zeven principes van ethici, rechtsgeleerden, bestuurskundigen en AI-experts. De doelen waren het opstellen van kaders voor AI-ontwikkeling en -toepassing, principes die iedereen van AI laten profiteren en het faciliteren van het debat over op gelijkheid gerichte, inclusieve en duurzame AI. Dit proces culmineerde in de Montreal Declaration of Responsible AI. Ook het eerdergenoemde AI Now Institute draagt bij aan het verbeteren van het gebruik van AI door de ontwikkeling van een *Algorithmic Accountability Policy Toolkit* en *Algorithmic Impact Assessments*.

Een andere prominente organisatie die zich inzet voor het verbeteren van AI, is de Partnership on AI. Leden ervan zijn grote bedrijven als Amazon, Facebook, Google, Deepmind, Microsoft en IBM, maar ook het Chinese Baidu. De Partnership is een non-profitorganisatie gericht op het verantwoordelijk gebruik van AI en doet dat door goede praktijken te identificeren en kennis te delen.[610] De organisatie ontwikkelde onder andere een database voor AI-incidenten waarin incidenten met zelfrijdende auto's of flashcrashes op de beurs geïnventariseerd worden.

Een laatste organisatie die we hier noemen, is OpenAI. Dat is deels een op winst gerichte onderneming die onder meer aan de wieg stond van GPT-3, het AI-programma dat het artikel in *The Guardian* schreef waarmee dit rapport opende. Voor een ander deel doet de organisatie ook non-profitonderzoek gericht op het ontwikkelen van 'vriendelijke AI'. OpenAI heeft aanzienlijke financiering ontvangen van Elon Musk en Microsoft.

Ondanks dat de meerderheid van de initiatieven bij deze vorm van engagement hun oorsprong in het buitenland heeft, kent ook Nederland een waardevol voorbeeld. Dat is de Nationale AI-Cursus die als volgt gemotiveerd wordt: "Kunstmatige Intelligentie (AI) neemt een steeds belangrijkere plek in het dagelijks leven in. Toch bestaan er nog veel misverstanden over deze technologie en de toepassingen ervan. Daarom is er nu de Nationale AI-Cursus, een

610 Russell 2019: 250.

breed toegankelijk programma voor iedereen die zich goed wil voorbereiden op de toekomst." Volgens de initiatiefnemers neemt AI een steeds belangrijkere plek in het dagelijks leven, maar bestaan er nog veel misverstanden over wat de technologie is en kan. Het doel van de cursus is om iedereen die meer wil weten over AI, daarvoor op een toegankelijke manier informatie aan te reiken. Andere interessante Nederlandse initiatieven zijn de AI Impact Assessment van ECP en de door Peter-Paul Verbeek ontwikkelde begeleidingsethiek die steeds meer gebruikt wordt. Het momentum groeit dus voor deze vorm van engagement en naarmate AI dieper in de samenleving ingebed raakt, zal het belang van die vorm ook toenemen.

Toe-eigenen: diversiteit in doelen en belangen

De laatste vorm van engagement is het toe-eigenen van AI. Dit is de meest symbiotische vorm van engagement. In tegenstelling tot verbeteren gaat het hierbij niet over bijdragen aan toekomstig goed en rechtmatig gebruik van AI, maar om het daadwerkelijk in de praktijk brengen van AI door partijen uit het maatschappelijk middenveld. Een nieuwe systeemtechnologie wordt doorgaans eerst in de praktijk gebracht door het bedrijfsleven en door overheden, omdat zij daar de middelen voor hebben. Veel partijen uit het maatschappelijk middenveld volgen pas later. Denk hierbij vooral aan groepen burgers en professionele organisaties. Voor het toe-eigenen van AI door deze laatste twee groepen bespreken we hier enkele initiatieven.

Verscheidene initiatieven zetten zich in voor bevolkingsgroepen die door AI benadeeld worden (zie paragraaf 6.2). Het kritisch volgen en toetsen van AI-gebruik door bedrijven en overheden voerde daarbij de boventoon. Een stap verder is dat AI ook zelf wordt ingezet om de belangen van die groepen te behartigen. Ruha Benjamin benadrukt het belang dat gemeenschappen technologie gaan gebruiken om de uitsluitende werking ervan tegen te gaan. Zij beschrijft de democratisering van data door een initiatief als DiscoTech ('*discovering technology*'), dat technologie toegankelijk maakt zodat groepen haar zich in de praktijk kunnen toe-eigenen.[611]

De groep Mijente omschrijft zichzelf als 'politieke thuisbasis' voor Latino-Amerikanen en Mexicanen en werkt onder andere aan het in kaart brengen van de relatie tussen AI en immigratie. Media Justice is een organisatie die opkomt voor mensen van kleur en mensen met lagere inkomens, en streeft naar een eerlijke economie, verbonden gemeenschappen, en een politiek landschap waarin deze groepen zichtbaar zijn, een stem hebben en macht. Om dit te bereiken is volgens de initiatiefnemers een omgeving van media en technologie nodig die

611 Benjamin 2020: 188-189.

echte gerechtigheid voedt. Inmiddels is er een veelheid aan organisaties gericht op de belangen van minderheden, zoals Women in AI, Black in AI – mede opgericht door de door Google ontslagen Timnit Gebru – en Queer in AI. Enkele van die organisaties, zoals Women in AI, zijn ook in Nederland actief.

In Nederland vindt toe-eigening onder meer plaats in AI-labs. Veel van die labs zijn gericht op de toepassing van AI door bedrijven of de overheid. Maar ook partijen uit het maatschappelijk middenveld raken betroken. Zo is in 2021 het Civic AI Lab opgericht, een samenwerking tussen de UvA, de VU en de gemeente Amsterdam. En bij de Universiteit van Tilburg werken wetenschappers samen met partners als Greenpeace, het World Food Programme en het Jeroen Bosch Ziekenhuis om AI te benutten bij het realiseren van maatschappelijke opgaven op het terrein van klimaat, voedselschaarste en gezondheidszorg.[612] Een laatste voorbeeld is het mede door rechtswetenschapper Maurits Barendrecht opgezette Hague Institute for Innovation of Law (HiiL), dat onder meer met de zogenoemde Justice Accelerator diverse projecten in vooral Afrika heeft geïnitieerd waarbij AI wordt benut.[613]

Er gebeurt dus al het een en ander op het gebied van toe-eigening in Nederland, alhoewel partijen uit het maatschappelijk middenveld nog maar mondjesmaat lijken aan te haken. Vooral de klassiekere belangenorganisaties op het terrein van bijvoorbeeld huurders, patiënten, consumenten of leraren lijken nog weinig actief te zijn. Deels heeft dat te maken met het feit dat we nog aan het begin staan van het proces van de inbedding van AI in onze samenleving. Dat groepen die opkomen voor bepaalde publieke waarden zich AI toe-eigenen en dat proces zo mede vormgeven, staat daarom nog in de kinderschoenen. Zij hebben vaak nog maar beperkt weet van AI, laat staan dat zij ideeën hebben over hoe AI voor hun doeleinden te gebruiken. Tegelijkertijd is het belangrijk dat zij niet achterblijven, omdat er met AI voor hun achterban ook veel te winnen is.

Tekstbox 6.7 – PublicSpaces

PublicSpaces is niet direct op AI gericht, maar is een mooi voorbeeld van toe-eigening van digitale technologie door het maatschappelijk middenveld. Het is een samenwerking van meer dan twintig partijen uit de publieke media, cultureel erfgoedsector, festivalsector, musea, onderwijs en gezondheidszorg. Deelnemers zijn onder andere BNNVARA, de Dutch Design Week, de EO, het Hollandfestival, de OBA, Waag en het Eye Filmmuseum.

612 Zie: Tilburg University, z.d.; Data Science Center Tilburg, z.d.
613 HiiL, z.d.

De coalitie is opgericht om het internet als een publieke ruimte te herontdekken en de beginselen ervan nieuw leven in te blazen. PublicSpaces ageert tegen de afhankelijkheid van grote techbedrijven als het gaat om onze communicatie, informatie en mediacirculatie, en wil toewerken naar een alternatief software-ecosysteem dat draait om publieke waarden in plaats van commerciële belangen.

In dat kader ontwikkelt de organisatie onder andere '*public badges*', kwaliteitskeurmerken voor de codering en tooling van websites en softwareapplicaties op grond van de waarden van PublicSpaces. Ook werkt ze aan de implementatie van open source-initiatieven.

Toe-eigening is niet alleen van belang voor gemeenschappen en specifieke bevolkingsgroepen maar ook voor beroepsgroepen. Dat is een enorm groot veld en er vinden door AI veranderingen plaats op werkvloeren in allerlei beroepen (zie hoofdstuk 5). Hier gaat het ons om de belangen en waarden die bepaalde bevolkingsgroepen belichamen zoals artsen, leraren en advocaten. Op basis van hun opleiding en werkervaring hebben zij een bepaalde expertise die in het gebruik van AI geborgd moet worden. Het is met andere woorden nodig dat zij zich AI toe-eigenen op een manier waardoor zij zorg kunnen dragen voor bijvoorbeeld goed onderwijs voor leerlingen, de gezondheid van patiënten of de rechten van hun cliënten.

Veel AI-toepassingen worden gepresenteerd als vervangingen van dit soort expertise. Robotrechters of medische algoritmes zouden klassieke beroepsgroepen overbodig maken. Dat is, zoals we zagen, een misvatting die past binnen een antagonistische relatie tussen AI en samenleving. Naarmate AI echter in de samenleving ingebed raakt, is juist een symbiotischer relatie nodig. Dat betekent het combineren van de expertise van deze beroepsgroepen met AI. Frank Pasquale laat zien dat AI niet expertise ondermijnt, maar in de huidige vorm vooral een benadrukking is van een bepaald type expertise ten opzichte van andere. De expertise van computerwetenschappers en economen staat centraal bij veel AI-toepassingen en die krijgt de overhand op andere vormen van expertise.[614] Daarmee wordt bij zo'n toepassing vaak maar een enkele en simpele maatstaf als uitgangspunt genomen. Allerlei beroepsgroepen hebben echter een complexiteit aan maatstaven, doelen, belangen en kennis in te brengen. Algoritmes die artikelen schrijven, kunnen misschien iets dat journalisten kunnen, maar vervangen niet het hele palet aan zaken waar een professionele

614 Pasquale 2020: 23.

journalist voor verantwoordelijk is, zoals het meenemen van verschillende perspectieven, het recht doen aan personen of het doen van diepteonderzoek.

In de eerste fase van de intrede van AI in de samenleving lag de nadruk op het revolutionaire karakter ervan. Gaandeweg groeide de zorgen over de banen die d technologie overbodig zou maken. In de volgende fase is het echter van belang dat allerlei beroepsgroepen zich AI toe-eigenen vanuit hun professionele verantwoordelijkheden. Beroepsgroepen hebben allerlei vormen van zelfregulering, en toezichthoudende instanties die licenties verstrekken en praktijken monitoren. Daarbij dient het gebruik van AI in het werkveld meegenomen te worden.[615] Alvorens dat kan gebeuren, is het nodig dat beroepsbeoefenaren zich de technologie eigen hebben gemaakt en zicht hebben op de manieren waarop die een bijdrage aan hun werk kan leveren.

Kernpunten – Samenwerking: verbeteren, toe-eigenen

- Bij *samenwerking* is de houding ten opzichte van AI symbiotisch. Samenwerking omvat de inzet om de technologie te verbeteren en om de technologie toe te eigenen om eigen doelen en waarden te realiseren.
- Bij het *verbeteren* van de technologie gaat het om mensen die werkzaam zijn op het gebied van AI of een aanverwante of andere relevante expertise hebben. Zij werken aan AI vanuit de overtuiging dat deze de samenleving zal verrijken en wenden hun expertise over het onderwerp aan om AI en het gebruik ervan te verbeteren. Concrete vormen zijn het opstellen van principes, het schrijven van open brieven, het ontwikkelen van instrumenten voor goed AI-gebruik (toolkits) of andersoortige publicaties.
- *Toe-eigening* vindt nog weinig plaats. AI wordt vooral in de praktijk gebracht door het bedrijfsleven en door overheden; partijen uit het maatschappelijk middenveld en beroepsgroepen haken maar mondjesmaat aan. Toe-eigening is onder meer belangrijk omdat deze partijen AI kunnen gebruiken om de uitsluitende werking ervan tegen te gaan of bij AI-gebruik de waarden te borgen die zij in hun beroepspraktijk belichamen.

615 Pasquale 2020: 88.

6.4 Tot slot

De opgave van engagement gaat over de vraag wie bij AI betrokken moet worden. Bij een nieuwe systeemtechnologie zijn bedrijven en overheden veelal de eersten die ervan gebruik maken. Daardoor hebben ze een sterke invloed op de ontwikkelingsrichting van een technologie. Gaandeweg raken ook partijen uit het maatschappelijk middenveld bij dit proces betrokken. Denk hierbij aan belangengroepen, kennisinstellingen, media en beroepsgroepen.

Engagement is belangrijk in elke samenleving en bij uitstek in een democratische samenleving als de Nederlandse. De belangen, waarden en kennis van een veelheid van maatschappelijke actoren is nodig om een nieuwe technologie op verantwoorde wijze in de samenleving in te bedden. Dat betekent dat hun stem gehoord moet worden, zowel wanneer zij geraakt worden door het gebruik van de technologie, als ook al in het ontwerpproces ervan. Uiteindelijk dienen zij haar zelf te kunnen gebruiken om hun eigen doelen te realiseren. Anders geformuleerd, vanuit hun ervaringen en kennis leveren partijen in het maatschappelijk middenveld waardevolle terugkoppeling over AI die meegenomen moet worden om de technologie goed in de samenleving in te bedden.

Op dit moment is die terugkoppeling nog maar beperkt georganiseerd. Daardoor ontwikkelen bedrijven en overheden allerlei AI-toepassingen terwijl er onvoldoende zicht is op de effecten die deze hebben op het leven van burgers en van specifieke doelgroepen, maar ook zonder daarbij de kennis en expertise te betrekken die die groepen daarbij in kunnen brengen. Leraren én leerlingen hebben een rol te spelen bij de ontwikkeling van AI in het onderwijs, artsen én patiënten hebben die rol in de zorg en ga zo maar door.

In dit hoofdstuk stond het engagement rondom AI centraal. Daarbij hebben we verschillende vormen onderscheiden binnen een spectrum van een antagonistische tot een symbiotische verhouding tot AI. Een aantal antagonistischer vormen van engagement is al sterk ontwikkeld, zoals het protesteren en controleren. Dat is belangrijk en moet blijven doorgaan om kwalijk gebruik van AI tegen te gaan. Controleren vervult bovendien een belangrijke signalerende functie waarmee het kan bijdragen aan de totstandkoming van kaders, normen en regulering. Datzelfde geldt voor uitstappen. Werknemers van technologiebedrijven zitten dicht op het vuur en kunnen daarmee vroeg problemen identificeren. De meest antagonistische vorm, vechten, wordt bij AI nog niet veel gebruikt, maar kan een sterk signaal uit de samenleving vormen.

Er zijn partijen die zich bezighouden met het neutraal agenderen van AI in het maatschappelijk leven. Op internationaal niveau bestaan er ook verschillende initiatieven gericht op het verbeteren van AI door hiervoor principes te ontwikkelen en kennis en ervaring te delen. Engagement in de vorm van toe-eigening

is van groot belang, omdat partijen uit het maatschappelijk middenveld, gemeenschappen en beroepsgroepen in het bijzonder daarmee in staat worden gesteld om AI te gebruiken op een manier die bij ze past en waardoor ze AI kunnen gebruiken om hun doelen en waarden kunnen realiseren. Zowel klassieke belangengroepen als beroepsgroepen zijn nog maar beperkt in staat om zich AI toe te eigenen.

Er zijn dus ontwikkelingen op de neutrale monitorende en de symbiotische vormen van engagement, maar die zijn, in tegenstelling tot de antagonistische vormen, nog maar zwak ontwikkeld. Veel gebeurt ook op internationaal niveau en de overheid heeft een taak om nationale vormen van dit engagement te stimuleren en zo het maatschappelijk middenveld beter bij de inbedding van AI te betrekken.

Een mooi voorbeeld van de wijze waarop dit engagement valt te stimuleren, en vooral ook waarop de expertise van belanghebbenden valt te versterken, is de eerder in dit hoofdstuk genoemde Handreiking voor Ondernemingsraden die de SER in 2018 heeft opgesteld. Deze handreiking besteedt niet alleen aandacht aan de effecten van technologische ontwikkelingen op het werk, maar ook aan de rechten van ondernemingsraden om zaken ter discussie te stellen en aan zaken waar zij alert op moeten zijn. Verder formuleert de SER concrete voorbeeldvragen om daarmee ondernemingsraden te equiperen om proactief te handelen. Dergelijk engagement blijft tot op heden echter beperkt tot spaarzame initiatieven zoals die van de SER. Juist met oog op de publieke waarden die in het geding (kunnen) zijn, is het van belang dat de overheid groepen van belanghebbenden equipeert bij constructief-kritische vormen van engagement.

7. Regulering

De inbedding van AI in de samenleving vraagt ook om kaders, en daarmee om regulering. Nu de technologie overgaat van het lab naar de samenleving, is er inmiddels breed aandacht voor de effecten die AI heeft op de economie en de samenleving. En daarmee is er tevens discussie over de noodzakelijke reguleringsmaatregelen voor een legitieme inbedding van AI binnen de samenleving en overheidsprocessen.[616]

Behalve voor de kansen, is er daarbij ook bijzondere aandacht voor de potentiële negatieve uitwerkingen van AI. Inmiddels zijn er honderden AI-richtlijnen, gedragscodes, private standaarden, publiek-private samenwerkingsvormen en certificeringsprogramma's verschenen om zowel de kansen te faciliteren als de risicovolle uitwerkingen te adresseren.[617] Belangrijk te midden van deze initiatieven is het voorstel dat de Europese Commissie in het voorjaar van 2021 presenteerde voor een Verordening inzake AI.[618] En natuurlijk zijn potentieel ook diverse bestaande wettelijke bepalingen en kaders van toepassing, variërend van fundamentele rechten tot aansprakelijkheid en intellectuele eigendomsrechten, archivering en bewijsvoering. Kortom, de effecten van AI worden intussen in banen geleid middels een breed scala aan kaders en specifieke regels en daar zullen de komende jaren nog diverse maatregelen aan worden toegevoegd.

Bij de zoektocht naar de gewenste en noodzakelijke regulering gaat het niet alleen om de inhoudelijke normering via concrete regels, maar ook om het instrumentarium waarin de normen worden neergelegd (wettelijke regels of private afspraken) en het niveau waarop de regels worden afgekondigd (internationaal, nationaal, decentraal). Kort samengevat, staat bij de opgave van regulering de vraag centraal: *Wat voor kaders zijn nodig*? Juist omdat het bij AI om een systeemtechnologie gaat, zijn bij het adresseren van deze vraag specifieke kwesties van belang. Die gaan niet uitsluitend over de toepasbaarheid van bestaande regels of de noodzaak van nieuwe, maar vooral over de reikwijdte van deze kaders en het niveau waarop ze geformuleerd moeten worden. In dit hoofdstuk bespreken wij regulering specifiek vanuit de rol en positie van de (nationale en internationale) overheid, meer in het bijzonder de wetgever. Bij deze discussie komt natuurlijk ook aan de orde in welke mate de overheid bij de regulering van AI kan en dient te leunen op de betrokkenheid van andere partijen. Denk aan de technologiebedrijven die de AI-toepassingen in de samenleving brengen.

---

616 Meijer et al. 2021.
617 Voor een overzicht zie: Jobin et al. 2019.
618 Europese Commissie 2021b.

De kwesties die spelen bij de regulering van AI, vallen uiteen in twee delen, die we in dit hoofdstuk afzonderlijk behandelen. Om te beginnen is er de wisselwerking tussen regulering enerzijds en de ruimte voor innovatie anderzijds. Juist omdat AI een systeemtechnologie is, zal de overheid op een groot aantal fronten moeten aftasten wat nodig is. Veel van de implicaties van de introductie van AI in de samenleving zijn immers nog onduidelijk en onzeker. Dit heeft zijn weerslag op de keuzes van de overheid ten aanzien van regulering en de effecten ervan. Er zijn daarbij verschillende vragen aan de orde. Bieden de bestaande rechtsregels bijvoorbeeld voldoende rechtszekerheid en rechtsbescherming? Faciliteren die rechtsregels innovatie in voldoende mate? En hoe om te gaan met het gegeven dat wetgeving vrijwel altijd achter de technologische ontwikkeling aanloopt?

Als nieuwe regels noodzakelijk blijken, speelt bovendien de vraag wat er dan nationaal en wat internationaal moeten gebeuren, waar marktpartijen zelf verantwoordelijkheid kunnen nemen en waar de overheid zaken dient op te pakken. Hoewel op Europees niveau de duidelijke keuze is gemaakt om specifiek voor AI een wetgevend kader te introduceren, resteren tegelijkertijd talloze – ook fundamentele – kwesties die niet door dit specifieke Europese AI-regime worden bestreken. Bijvoorbeeld hoe algoritmische besluitvorming in relatie tot de rechtspositie van burgers de verschillende constitutionele uitgangspunten – waaronder het legaliteitsbeginsel – onder druk zet of zelfs transformeert.[619] Of welke implicaties de nieuwe EU-Richtlijn auteursrecht heeft voor de toegang tot data waarmee AI-systemen getraind worden.[620] En de talloze andere vragen die spelen in relatie tot data die worden gebruikt bij AI.[621] Evenzeer spelen vragen over mededinging en marktfalen[622] op het terrein van digitale diensten, bestuursrechtelijke implicaties, de noodzaak tot archivering van algoritmes gegeven de vereisten van de Archiefwet[623] en zelfs de kwestie of onze constitutie is opgewassen tegen AI.[624] Kortom, het Europese AI-regime is een belangrijke stap, maar de wetgever zal aanvullend nog talloze vragen moeten beantwoorden. In het eerste deel van dit hoofdstuk staan we daarom stil bij diverse generieke kwesties die spelen rond het instrumentarium om AI te reguleren, meer in het bijzonder het instrument wetgeving. We richten de blik hier dus niet op de concrete juridische vragen die op tal van terreinen voorliggen, maar op de

619 Goossens et al. 2021.
620 EU-Richtlijn 2019/790 van 17 april 2019 inzake auteursrechten en naburige rechten in de digitale eengemaakte markt (L130/92).
621 CPB 2021.
622 CPB 2019.
623 Helwig 2020.
624 Passchier 2020.

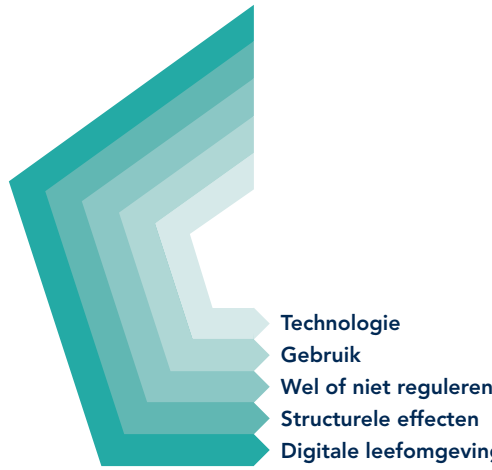
overwegingen die relevant zijn bij de wijze waarop de wetgever deze concrete vragen kan adresseren.

Vanuit de historische lessen die zijn te trekken over de omgang van de overheid met eerdere systeemtechnologieën, wordt tevens duidelijk dat het proces van inbedding van een technologie telkens weer gepaard gaat met een grotere rol van de overheid. Dat is het vertrekpunt van het tweede deel van dit hoofdstuk, waarbij het onder meer gaat over de invloed van de factor tijd op de kaders en de concrete spelregels voor AI. Gedurende het proces van maatschappelijke inbedding geven rechtspraak, toezichthouders, ngo's en parlement signalen af over het vermogen van de samenleving dan wel de publieke sector om zelf, dus zonder interventie van de wetgever, de inbedding van de technologie in de samenleving in te kaderen. Het gaat dan bijvoorbeeld om het vermogen van de samenleving en overheidsinstanties om bij de toepassing van een systeemtechnologie ook publieke waarden voldoende in acht te nemen. Veel van deze signalen agenderen in de praktijk de noodzaak tot interventie door de wetgever, rechtspraak of toezichthouders.

Bij vrijwel iedere systeemtechnologie blijkt in de tijd een steeds grotere invloed van de overheid te worden gevraagd, en daarmee een explicietere rol voor wetgeving. Het gebruik van stoommachines bijvoorbeeld leidde tot zeer schadelijke luchtverontreiniging in steden, die pas afnam nadat ondernemingen verplicht werden hogere schoorstenen te gebruiken.[625] Deze, ook bij AI te verwachten, groeiende rol van de overheid agendeert de vraag hoe de overheid zich hierop kan voorbereiden. We betogen dat het daarvoor in ieder geval nodig is de blik te verbreden van een perspectief dat momenteel nog primair lijkt te zijn gericht op de technologie zelf, naar een perspectief dat tevens het proces en de effecten van inbedding in de samenleving omvat. Naast regulering die vooral toeziet op de ontwikkeling, de kenmerken en het gebruik van AI, moet er dus ook regulering komen voor de effecten van de maatschappelijke inbedding van AI. Ook laten we zien dat naarmate de rol van AI in onze samenleving groter wordt, er vaker fundamentele keuzes aan de orde zijn over de inrichting van wat we in dit hoofdstuk 'de digitale leefomgeving' noemen. Concreet betekent dit dat het debat over regulering zich de komende jaren niet kan beperken tot kwesties als betrouwbaarheid, transparantie, en privacy maar zich tevens dient te richten op de bredere vraagstukken over de inrichting van een samenleving waarin AI een prominente plaats inneemt (zie figuur 7.1). Het gaat dan in ieder geval over partijen die het voor het zeggen hebben wat betreft de inrichting van de digitale leefomgeving alsmede het instrumentarium dat zij daartoe benutten. Bij dit laatste gaat het in het bijzonder over het instrument 'data'.

625 Bakker en Korsten 2021: 58.

Figuur 7.1 Verschillende niveaus van regulering



7.1 Normering van AI door de overheid

Met de presentatie van het concept voor een AI-Verordening geeft Europa een helder signaal af over de noodzaak van specifieke regels voor AI. Tegelijkertijd resteren daarmee nog talloze kwesties die de verordening niet afdekt, of die in aanloop naar de vaststelling daarvan nog verheldering behoeven. Daarmee blijft de vraag relevant of bestaande kaders al dan niet toepasbaar zijn op AI. Hoe zit het bijvoorbeeld met een eventueel noodzakelijke aanpassing van de Algemene wet bestuursrecht of de talloze regelingen die gelden voor specifieke sectoren, zoals de zorg en mobiliteit?

Over deze en andere kwesties is de afgelopen jaren het nodige geschreven en gedebatteerd en regelmatig resulteerde dat ook in noodzakelijke aanpassingen van deze kaders.[626] Op deze plaats gaan wij niet in op dit debat en de betreffende aanpassingen. We richten ons hier namelijk op de specifieke vragen rond het instrumentarium (type, niveau), in het bijzonder op de vragen die naar voren komen gegeven het systeemkarakter van AI.

Om te beginnen zijn dat vragen die verband houden met de breedte van de impact die AI zal hebben. Als systeemtechnologie heeft AI de potentie om alomtegenwoordig te worden en op veel terreinen complementaire innovaties in gang te zetten. De technologie valt met andere woorden in te zetten in een

626 Voor een recent overzicht zie: Fierens et al. 2021; Van Gool et al. 2021; Chavannes et al. 2021.

zeer groot aantal domeinen en bovendien voor een zeer groot aantal doeleinden. Het besef hiervan is groeiende, zoals onder andere blijkt uit de in hoofdstuk 2 besproken onderzoeken naar AI in sectoren als de zorg, het onderwijs en defensie. Als gevolg hiervan rijzen vragen over de (wettelijke) voorwaarden waaraan de ontwikkeling en het gebruik van AI in die specifieke sectoren moeten voldoen. Gegeven het systeemkarakter van AI is het zaak bij de discussie over de noodzakelijke regulering te kijken of de regels moeten worden toegesneden op de specifieke sector, het type AI-techniek dat in gebruik is of de concrete toepassing ervan. Wellicht kan in sommige gevallen ook worden volstaan met generieke regels, waaronder regels die een bredere gelding hebben dan specifiek voor AI.

In het nadenken hierover is het onderscheid behulpzaam dat Lyria Bennett Moses maakt in haar onderzoek naar regulering en technologische verandering. Zij werkt met vier categorieën vraagstukken die een nieuwe technologie kan oproepen.[627] Allereerst kan het nodig blijken om nieuwe handelswijzen te reguleren. Dit zien we bijvoorbeeld bij de bepaling in de Algemene Verordening Gegevensbescherming (AVG) dat er een mens aan te pas dient te komen wanneer een autonoom handelend systeem persoonsgegevens verwerkt en het hierbij gaat om de rechtspositie van personen. De tweede categorie die Bennett Moses onderscheidt, is de noodzaak om bestaande regels te verhelderen, bijvoorbeeld omdat onduidelijk is of ze van toepassing zijn. Diverse bestaande regels zijn destijds immers niet met het oog op AI opgesteld, wat de ontwikkeling van nieuwe en voor de samenleving relevante toepassingen van de technologie ten onrechte kan blokkeren. Illustratief voor AI is hier de recente discussie op EU-niveau over de vraag aan welke partij in juridische zin de rechtshandelingen, uitgevoerd door AI, toegerekend dienen te worden.[628] En zou die partij ook het systeem zelf kunnen zijn?[629] Deze laatste vraag is onderdeel van de bredere discussie over de vraag of rechten en plichten aan AI-systemen toegekend moeten worden.[630] Ook eerdere systeemtechnologieën kenden een fase waarin de al dan niet toepasselijkheid van bestaande regels verhelderd diende te worden. Zo had in ons land de Hoge Raad in 1921 te oordelen over de vraag of elektriciteit gestolen kon worden. Bij elektriciteit kon vanwege het immateriële karakter ervan namelijk geen sprake zijn van 'wegnemen van een goed'.[631]

---

627 Bennett Moses 2007a.
628 Zie het onderzoek naar burgerlijke aansprakelijkheid dat Bertolini (2020) uitvoerde in opdracht van de Commissie Juridische Zaken van het Europese Parlement, en de resolutie van het Europese Parlement over datzelfde onderwerp (Europees Parlement, 20 oktober 2020).
629 Hage 2017.
630 Brown 2021.
631 Elektriciteitsarrest (Hoge Raad, 23 mei 1921).

De derde categorie van Bennett Moses betreft de noodzaak tot regulering omdat risicovolle toepassingen ongecontroleerd in gebruik worden genomen. Ook van deze categorie zien we voorbeelden in de historie. Zo leidde de toenemende dichtheid van stroomnetwerken in stedelijke centra tegen het einde van de negentiende eeuw tot een warboel van stroomkabels en het probleem van *death by wire* als gevolg van onveilig geïnstalleerde kabels. Dit werd opgelost door regulering die de private nutsbedrijven ertoe verplichtte verantwoordelijkheid te nemen voor de veiligheid van het elektriciteitsnet. Illustratief voor AI is de Europese concept-Verordening inzake AI, waarmee de EU specifieke applicaties verbiedt, zoals toepassingen die kwetsbare personen uitbuiten, gericht zijn op willekeurige massasurveillance voor rechtshandhaving, zien op *social scoring* (zoals het welbekende socialekredietsysteem van de Chinese regering) en schadelijke manipulatie. Een voorbeeld uit eigen land is de oproep van de Tweede Kamer om "te stoppen met het gebruik van discriminerende algoritmes". Deze oproep is onderdeel van een meeromvattende motie die in het voorjaar van 2011 met een ruime meerderheid van stemmen werd aangenomen als uitvloeisel van de parlementaire discussie over de notulen van de ministerraad in de Toeslagenaffaire.[632]

Tot slot – de vierde door Bennett Moses benoemde categorie – kan een nieuwe technologie vragen oproepen omdat de bestaande regels gebaseerd zijn op aannames die niet langer geldig zijn, waardoor het dus niet langer gerechtvaardigd is om de betreffende regel conform deze aanname te handhaven. Een voorbeeld in de recente geschiedenis is de verbreding van het beschermingsoogmerk bij de strafbaarstelling van kinderporno. Voor de komst van digitale technieken beoogde de regeling in art. 240b Sr. te voorkomen dat kinderen werden misbruikt om kinderporno te fabriceren. Maar met de mogelijkheden van digitale technieken om afbeeldingen te manipuleren was feitelijk misbruik niet langer noodzakelijk bij het maken van de afbeeldingen. Bij wetswijziging in 2002 werd het oogmerk daarom zodanig aangepast dat onder de bepaling ook afbeeldingen vallen die schadelijk voor kinderen zijn zonder dat voor het maken daarvan een gedraging noodzakelijk is.

De overheid zal de komende jaren voor een zeer groot aantal maatschappelijke domeinen, en daarmee ook rechtsdomeinen, moeten nagaan in hoeverre de regels die daarbinnen gelden, voldoende zijn toegesneden op nieuwe entiteiten, activiteiten en relaties die ontstaan door AI. Het geschetste onderscheid in de vier categorieën kan daarbij behulpzaam zijn. Indien de constatering vervolgens is dat (aanpassing van) regulering aan de orde is, komen diverse vervolgvragen in beeld. De eerste vraag is of regulering specifiek of juist generiek dient te zijn.

632 Kamerstukken II 2020/21, 28362, nr. 44.

De tweede vraag is of een technologieneutrale of een meer toegespitste benadering wenselijk is. De derde en vierde vraag gaan over het niveau van de regulering en de actoren en het instrumentarium dat deze kunnen inzetten. Deze vragen behandelen we hieronder kort, waarbij we specifiek ingaan op kwesties die spelen nu AI als een systeemtechnologie aangemerkt dient te worden.

Specifiek of generiek beleid

De alomtegenwoordigheid van AI kan het idee oproepen dat de regulering ervan vraagt om generieke kaders. Een voorbeeld van deze gedachtegang zien we in zowel de discussie over transparantie en uitlegbaarheid als die over de instelling van nieuwe AI-autoriteiten. Toch lijkt een generieke aanpak op termijn nauwelijks realistisch. We lichten dit kort toe.

In het debat over regulering van AI is er volop aandacht voor transparantie en uitlegbaarheid.[633] Niet alleen de technische werking van het systeem, waaronder de beslisregels, moet uiteen worden gezet op een manier die voor mensen te begrijpen valt. Ook moeten de keuzes die aan de inzet van de AI-technieken ten grondslag liggen en de feitelijke beslissing die het systeem neemt, verhelderd kunnen worden.[634] Het belang hiervan is onder meer ingegeven door de noodzaak om aan te kunnen tonen of de rechtsbescherming van burgers dan wel fundamentele rechten in het gedrang komen.[635] Transparantie en uitlegbaarheid zijn ook belangrijk met het oog op de vaststelling van de aansprakelijkheid en de verantwoordelijkheid voor door het AI-systeem genomen beslissingen. Zeker als duidelijk moet worden hoe een AI-systeem redeneert en waarom het tot bepaalde beslissingen is gekomen. Verhoogde transparantie over de werking van algoritmes kan echter ook concurrenten in de kaart spelen doordat bedrijfsmodellen dan openbaar gemaakt worden. Zoals Gerbrandy en Custers opmerken, kan transparantie de mededinging beperken én haaks staan op de bescherming van intellectuele-eigendomsrechten.[636]

Bovendien: wat precies behelzen transparantie en uitlegbaarheid? Beide kunnen immers niet alleen veel betekenen, maar ook vanwege verschillende doeleinden worden nagestreefd. De uiteindelijke keuzes die in dit verband worden gemaakt, zijn van grote invloed op het type informatie dat beschikbaar wordt gemaakt en aan wie. Illustratief is een onderzoek, uitgevoerd in samenwerking

633 Zie onder meer: Kamerstukken II 2018/19, 26643, nr. 570, p. 3-4; Kamerstukken 2019/20, 26643 en 32761, nr. 641.
634 Zie in dit verband over de besluitvorming binnen het openbaar bestuur: Coglianese en Lehr 2019
635 Van Eck et al. 2018. Zie ook het Jaarverslag 2020 van de Raad van State, waarin bijzondere aandacht wordt gevraagd voor de risico's op stigmatisering, stereotypering en discriminatie (Raad van State 2020: 42).
636 Gerbrandy en Custers 2018: 108.

met het CBS, waaruit blijkt dat wetenschappers bij uitlegbaarheid vooral bedoelen dat ze het aan elkaar kunnen uitleggen, niet aan de burger.[637] Soms ook zal de context waarin AI wordt toegepast, impliceren dat bij de concretisering van transparantie en uitlegbaarheid specifieke eisen gelden. Bijvoorbeeld wanneer AI wordt ingezet bij medische diagnoses en de daarop gebaseerde behandeling. Het relatief streng ingevulde vereiste van toestemming van de patiënt – 'informed consent' – heeft dan ook betekenis voor de mate van detail waarin de arts uitleg moet geven over hoe de 'advisering' door het AI-systeem tot stand kwam.[638] Illustratief voor de discussie over generieke dan wel specifieke regels is ten slotte ook de vraag of transparantie vooraf of enkel achteraf een vereiste zou moeten zijn. Het belang daarvan verschilt per toepassing en hangt tevens af van de ernst van de gevolgen die kunnen optreden. Transparantie vooraf kan het gebruik van sommige AI-toepassingen, zoals neurale netwerken, beperken. Christopher Reed betoogt daarom dat de eis van ex-antetransparantie beperkt zou moeten blijven tot situaties waarin AI risico's met zich meebrengt voor fundamentele rechten, of situaties waarin de samenleving verzekerd moet worden dat AI veilig te gebruiken is.[639]

Kortom, alhoewel voor transparantie en uitlegbaarheid op het eerste oog een generieke regel lijkt te kunnen volstaan, blijkt het in de praktijk toch vaak nodig de regeling specifiek te maken. Illustratief in dit verband zijn de uitspraken van de Raad van State in de AERIUS-zaak (2017 en 2018). Zie tekstbox 7.1.

Tekstbox 7.1 – AERIUS

De AERIUS-zaak uit 2017 ging over een landelijk computerprogramma dat de Programmatische Aanpak Stikstof van de Nederlandse overheid ondersteunt en waarvan lokale overheden gebruik moeten maken bij de vergunningsverlening voor stikstofverhogende activiteit. In een zaak over dit programma formuleerde de Afdeling Bestuursrechtspraak van de Raad van State (ABRVS) voor het eerst een duidelijk toetsingskader waaraan de overheid moet voldoen als zij het nemen van besluiten gedeeltelijk of helemaal overlaat aan een computerprogramma.[640]

De ABRVS concludeerde dat een gebrek aan transparantie over de gemaakte keuzes en gebruikte gegevens en aannames het voor rechtszoekenden onmogelijk maakt om het geautomatiseerd genomen

637 De Ree, 29 april 2021.
638 Klincewicz en Lily 2020.
639 Reed 2018.
640 Afdeling Bestuursrechtspraak van de Raad van State, 17 mei 2017.

overheidsbesluit ter discussie te stellen. De ABRVS formuleerde daarom een toetsingskader dat niet alleen ziet op transparantie en uitlegbaarheid van de AI-techniek, maar ook van de datasets aan de hand waarvan het systeem werd getraind en de wijze waarop het leerproces van het systeem vorm kreeg.

Opvallend is dat de ABRVS in de tweede uitspraak over deze kwestie de vereisten voor transparantie die hij in eerste instantie geformuleerd had, nader preciseerde en wel door een onderscheid te maken tussen standaard- en maatwerkinvoergegevens.[641] Van burgers mag worden verwacht dat zij zelf tijdig om inzage in het eerste type gegevens vragen, aldus de bestuursrechter. Kortom, deze gegevens vormen geen onderdeel van de vereiste transparantie bij de AI-toepassing. En daarmee is het toetsingskader specifieker geworden. Inmiddels hebben ook de Hoge Raad en de Centrale Raad van Beroep het toetsingskader van de ABRVS toegepast in een zaak over de WOZ-waardebepaling van huizen, die gemeenten eveneens met behulp van speciale software vaststellen.[642]

Ook uit andere praktijkvoorbeelden komt naar voren dat zowel de vereisten als de feitelijke mogelijkheden voor transparantie en uitlegbaarheid van tal van factoren afhangen.

Zo laat onderzoek van Stefan Kulk en Stijn van Deursen, in opdracht van het WODC, zien dat kansen en risico's voor publieke waarden sterk afhankelijk zijn van de domeinen en de organisatorische context waarin een algoritme wordt ingezet.[643] Dat een CJIB-medewerker een schuldenaar onterecht belt, is volgens de auteurs van een andere orde dan een verkeersongeluk dat een zelfrijdende auto veroorzaakt door een verkeerde interpretatie in de sensoriek van de auto. Ook het modereerproces van onlineplatformen, waarin algoritmes redengevend zijn maar niet zelf beslissen, valt volgens hen niet te vergelijken met algoritmes die rechterlijke uitspraken automatisch kunnen anonimiseren. Dergelijke verschillen in domeinen, de organisatorische context en de verhoudingen tussen de betrokken partijen, pleiten er volgens Kulk en Van Deursen voor om eventuele knelpunten zo veel mogelijk domeinspecifiek aan te pakken.

641 Afdeling Bestuursrechtspraak van de Raad van State, 18 juli 2018.
642 Hoge Raad, 17 augustus 2018; Rechtbank Amsterdam, 4 juli 2019; Centrale Raad van Beroep, 15 mei 2019.
643 Kulk en Van Deursen 2020.

De kwestie van specifieke of generieke regulering speelt, zoals gezegd, ook bij de discussie over het toezicht op AI en de voorstellen voor een nieuwe algemene AI-toezichthouder of AI-autoriteit, bijvoorbeeld in Nederland, de EU en de VS.[644] Ook hier geldt dat een generieke aanpak – een algemene toezichthouder die over specifieke deskundigheid beschikt – aantrekkelijk oogt. Toch zijn er verschillende argumenten die tegen een dergelijke aanpak pleiten. Een toezichthoudende instantie heeft een afgebakend werkveld nodig en een set overkoepelende principes als basis voor het toezichtsregime. Zoals ook bij andere technologieën die zich in de prille fase van toepassing bevinden, is het op dit moment niet mogelijk om dergelijke principes op te stellen omdat de risico's nog onvoldoende bekend zijn.[645] Maar waarin AI verschilt met niet-systeemtechnologieën, is de vraag of een regime valt te ontwerpen dat moet toezien op alle mogelijke AI-toepassingen. Dat regime dient namelijk een ongelofelijke reikwijdte te hebben en moet tegelijkertijd een enorme variëteit van vraagstukken kunnen adresseren. De toepassingen van AI zijn immers zeer divers en soms onvergelijkbaar in hun implicaties. Het zou onwenselijk zijn om het regime voor autonome voertuigen ook toe te passen op slimme koelkasten die, gebaseerd op consumptiepatronen, levensmiddelen bestellen. Het probleem is hier dus niet toezichtbeleid als zodanig, maar te generiek beleid dat vervolgens weer talloze uitzonderingen zal kennen in het licht van specifieke applicaties. Het potentieel enorme mandaat van een algemene AI-autoriteit of AI-toezichthouder is bovendien problematisch in het licht van de constatering, door Verhey en Verheij, dat de bestaande rechtsbescherming bij toezichtshandelingen sowieso verbetering behoeft gegeven de ingrijpende handhavingsbevoegdheden waarover zij beschikken.[646] En recenter deed de Raad van State in het ongevraagde advies over ministeriële verantwoordelijkheid juist de aanbeveling om terughoudend te zijn met het neerleggen van allerlei taken bij onafhankelijke instellingen (waaronder toezichthoudende instellingen) omdat hierdoor de mogelijkheid van democratische controle te veel wordt beperkt.[647]

De vraag of AI om generieke of juist specifieke kaders vraagt, keert ook terug in het concept voor de Verordening inzake AI van de Europese Commissie, zowel met betrekking tot de risicobenadering als met betrekking tot het bijbehorende toezicht (zie tekstbox 7.2). De Europese Commissie snijdt met deze verordening de regulering van AI zo veel mogelijk toe op concrete situaties waarin AI nu of in de nabije toekomst tot risico's kan leiden. Tegelijkertijd bevat de verordening

644 Zie de diverse bijdragen in de special van het *Tijdschrift voor Toezicht* (nr. 1, 2020) over voorspellende modellen, algoritmes en AI.
645 Nemitz (2018) noemt desondanks een paar richtingen waarin we die principes zouden kunnen zoeken.
646 Zie Verhey en Verheij 2005.
647 Raad van State 2020.

ook flexibele mechanismen, die aangepast kunnen worden naarmate AI zich verder ontwikkelt en er nieuwe risico's ontstaan.

Tekstbox 7.2 – Algemene en specifieke kaders in de concept-Verordening AI

De Europese Commissie onderscheidt vier risicocategorieën, gekoppeld aan specifieke AI-technieken, doeleinden en sectoren. Zij geeft hiermee het signaal af dat AI-technieken en -toepassingen niet over één kam geschoren kunnen worden en niet allemaal dezelfde impact op de samenleving hebben. De Commissie kiest zodoende voor een specifieke benadering van AI. De vraag is nu welke toepassingen onder welk risicoregime moeten gaan vallen. Zo gaat het verbod op biometrische identificatie verschillende partijen niet ver genoeg[648] en is er de vraag waarom het verbod op *social scoring* uitsluitend publieke organisaties betreft, terwijl juist de private sector intensief betrokken is bij de gedataficeerde verzorgingsstaat.[649] Hiernaast speelt het duale gebruikskarakter van bepaalde AI-applicaties en het feit dat aanbieders systemen zo kunnen ontwerpen dat afnemers ze nadien voor manipulatieve doeleinden kunnen aanpassen. Dat de Europese Commissie de verkoop van AI-systemen met een manipulatief oogmerk aan repressieve regimes verbiedt, is daarmee relatief eenvoudig te omzeilen. Een fundamenteel kritiekpunt is tot slot dat de Commissie onvoldoende oog heeft voor het onrecht en de (immateriële) schade die AI-systemen in termen van fundamentele rechten kunnen veroorzaken, en deze systemen daardoor aan een te lichte toets zou onderwerpen.[650]

De kwestie van algemeen versus specifiek beleid zien we eveneens gereflecteerd in de keuze voor het governancesysteem, dat overwegend moet voortbouwen op bestaande structuren in de lidstaten. Zo stelt de Europese Commissie in de verordening voor dat elke lidstaat één of meerdere nationale autoriteiten of toezichthouder(s) aanwijst, waar mede de verantwoordelijkheid ligt om de verordening uit te werken en na te leven. Aanvullend stelt de Commissie voor om, afhankelijk van de sector waar een AI-systeem in gebruik wordt genomen, ook toezichthouders aan te wijzen voor die specifieke sector.

---

648 Bijvoorbeeld European Data Protection Board en European Data Protection Supervisor 2021.
649 Chiusi et al. 2020.
650 Smuha et al. 2021.

Juist het kenmerk van de alomtegenwoordigheid van AI maakt dus dat de wettelijke vereisten verschillende aspecten moeten kunnen verdisconteren en derhalve niet per definitie generiek kunnen zijn. Toch zijn generieke kaders wel degelijk relevant, zeker wanneer het de overheid is die AI toepast. De Raad van State wijst in dit verband op de voor de positie van burgers zo belangrijke waarborg van eenheid van wetgeving.[651] En ten aanzien van concrete generieke kaders wijzen zowel de Raad van State als bestuursrechtjuristen bijvoorbeeld op de rol van de algemene beginselen van behoorlijk bestuur.[652] Wolswinkel ziet de door de minister van Rechtsbescherming opgestelde Richtlijnen voor het toepassen van algoritmes door overheden dan ook als een 'rechtstreeks gevolg' van deze algemene beginselen.[653] Specifieker op digitalisering toegeschreven, maar wel generiek, zijn de door Franken in de jaren negentig van de vorige eeuw ontwikkelde Algemene Beginselen van Behoorlijk ICT-gebruik.[654] Bijna dertig jaar na dato kunnen deze beginselen – beschikbaarheid, vertrouwelijkheid, integriteit, authenticiteit, flexibiliteit en transparantie – ook bij AI sturing geven bij het zoeken naar de goede balans tussen enerzijds een effectieve waarborg van publieke waarden en anderzijds voldoende ruimte voor nieuwe ontwikkelingen in de technologie.

De conclusie is derhalve dat de afweging en keuze tussen generiek en specifiek nooit zwart-wit zal zijn. Wel is een gedegen discussie hierover relevant, maar dan vanuit het besef dat het opstellen van kaders bovenal een zoektocht is naar een goede balans tussen enerzijds een effectieve waarborg van publieke waarden en anderzijds voldoende ruimte voor innovatie. Bovendien kan een te grote gerichtheid op specifieke sectoren ertoe leiden dat het algemene beeld ondergesneeuwd raakt, omdat AI gezien haar systeemkenmerken nu eenmaal niet te vangen is een bepaald beleidsterrein of een bepaald gebied van wetgeving.[655] Het is vanuit deze opdracht dat de overheid telkens weer de discussie zal moeten voeren over de vraag hoe specifiek of abstract de kaders voor AI moeten zijn.

Technologiespecifieke en technologieneutrale regels

Een tweede kwestie die speelt bij de regulering van AI, betreft de mate waarin wettelijke regels neutraal moeten zijn of juist specifiek dienen te zijn toegeschreven op de kenmerken van AI. Technologieneutrale wetgeving heeft verschillende voordelen.[656] Ten eerste zijn regels daardoor generiek en efficiënt toe te passen in verschillende technologische contexten. Ten tweede raakt een

651 Raad van State 2021: 115.
652 Raad van State 2021: 105-108.
653 Wolswinkel 2020.
654 Franken 1993.
655 Black en Murray 2019.
656 Voor een kritische bespreking zie Bennett Moses 2007b en Koops 2006.

technologieneutraal geformuleerde wet of bepaling minder snel verouderd wanneer de technologie verandert. Het idee hierachter is dat het bij een dergelijk type wetgeving gemakkelijker is te deduceren hoe deze moet worden toegepast. Het is dan namelijk mogelijk om terug te grijpen op meer algemene principes. Technologieneutrale wetgeving kan, om die reden, als een toekomstbestendiger vorm van wetgeving worden gezien en lijkt daarmee ook een geschikt instrument als het gaat om het opstellen van kaders voor AI.

Toch is ook deze keuze – juist vanwege het systeemkarakter van AI – niet zo eenvoudig als hij lijkt. Technologieneutrale wetgeving veronderstelt allereerst dat we een goed beeld hebben van hoe, min of meer functioneel equivalente, technologieën werken. Zo kan de wetgever bijvoorbeeld eisen stellen aan de remweg van auto's, zonder te specificeren welk remsysteem daarvoor moet worden gebruikt. Bij nieuwe technologieën is dit echter niet goed mogelijk, omdat de eigenschappen daarvan nog onbekend zijn. Ook kan een nieuwe technologie kwaliteiten bezitten die vragen om een andere afweging tussen wetgevingsdoelen. In dit verband wordt vaak gewezen op de spanning tussen accuraatheid en uitlegbaarheid. Vinden we uitlegbaarheid belangrijker dan accuraatheid, dan bevoordelen we regelgebaseerde AI-systemen boven systemen die van *deep learning* gebruik maken. Ten slotte kan een nieuwe technologie heel andere oplossingen in het vizier brengen om te voldoen aan de generieke doelstellingen van een wet, zoals in het voorbeeld van de verkeersveiligheid van medeweggebruikers. Als auto's op termijn bijvoorbeeld ook zouden kunnen gaan vliegen, dan wordt de hele idee van een remweg wellicht obsoleet. De Europese Commissie heeft – waarschijnlijk mede vanwege dit soort overwegingen – in de concept-Verordening inzake AI gekozen voor een functionele definitie van AI, gekoppeld aan een concrete lijst van technieken die nadien kan worden aangepast (tekstbox 7.3).

Tekstbox 7.3 – Technologiespecifieke en technologieneutreale wetgeving en de concept-Verordening AI

De Europese Commissie wil met de concept-Verordening voor AI een toekomstbestendig regime presenteren. AI wordt in het voorstel gedefinieerd als "software die is ontwikkeld aan de hand van een of meer van de technieken en benaderingen (…) die voor een bepaalde reeks door mensen gedefinieerde doelstellingen output kan genereren, zoals inhoud, voorspellingen, aanbevelingen of beslissingen die van invloed zijn op de omgeving waarmee wordt geïnterageerd." Door de functie van AI-systemen centraal te stellen en de technologie zelf niet te definiëren, beoogt de Commissie het wetgevend kader niet te hoeven aanpassen op het moment dat zich nieuwe ontwikkelingen voordoen.

Tegelijkertijd bevat de verordening wel een bijlage met een beschrijving van technieken en benaderingen die binnen de reikwijdte van de verordening vallen.[657] Een kritiek op deze aanpak van de Commissie is dat de reikwijdte van de verordening als geheel daarmee veel breder is dan de daarin opgenomen verplichtingen, die slechts betrekking hebben op kleinere onderdelen hiervan.

Relevant is ook dat AI weliswaar een systeemtechnologie is, maar in meerdere opzichten een uniek karakter heeft ten opzichte van eerdere systeemtechnologieën. In deel 1 spraken we in dit verband over AI-toepassingen als 'halffabrikaten', die naar hun aard constant veranderen. Ook grijpt AI vooral in op technologieën (computer, communicatiesystemen) die nu al vaak zonder menselijke tussenkomst functioneren. Daarnaast gaat het om een technologie die als het ware 'verdwijnt' in reguliere processen van het maatschappelijk leven. Juist deze kenmerken roepen specifieke vragen op over autonomie, aansprakelijkheid en verantwoordelijkheid. En daarmee over de noodzaak om AI juist indachtig deze kenmerken te reguleren en dus tot technologiespecifieke regulering te komen.

Onherroepelijk resulteert technologiespecifieke regulering in een 'gat' met AI-innovatie, in de zin dat wetgeving verouderd raakt en aangepast moet worden, hetgeen tijd kost, terwijl de innovatie doorgaat. Belangrijk is echter te constateren dat er inmiddels veel waardevolle kennis bestaat om met deze uitdaging om te gaan.[658] Ook moeten we vooral niet in de valkuil lopen van het idee dat technologische innovatie enerzijds en wetgeving anderzijds ten alle tijden gelijk op moeten lopen. Of zoals voormalig opperrechter van de Amerikaanse Supreme Court, Warren Berger, het een halve eeuw geleden al verwoordde: *"It should be understood that it is not the role and function of the law to keep fully in pace with science"*.[659]

---

657 Op deze lijst staan: a) benaderingen voor machinaal leren, waaronder gecontroleerd, ongecontroleerd en versterkend leren, met behulp van een brede waaier aan methoden, waaronder deep learning; b) op logica en op kennis gebaseerde benaderingen, waaronder kennisrepresentatie, inductief (logisch) programmeren, kennisbanken, inferentie- en deductiemachines, (symbolisch) redeneren en expertsystemen; c) statistische benaderingen, Bayesiaanse schattings-, zoek- en optimalisatiemethoden.

658 Zie bijvoorbeeld over het belang van een nadere typering van het gebrek aan aansluiting tussen wetgeving en technologische innovatie: Brownsword en Goodwin 2012; Bennett Moses en Gollan 2015.

659 Geciteerd in Marchant 2011: 27.

Van geval tot geval zal bepaald moeten worden waar wetgeving technologie-specifiek dan wel technologieneutraal dient te zijn.[660] En ook hier toont het verleden dat dit de min of meer natuurlijke gang van zaken is. De verplichte aansprakelijkheidswetgeving geldt bijvoorbeeld heel specifiek voor auto's, gegeven de omvang van de schade die kan voortvloeien uit auto-ongelukken. Deze wetgeving geldt daarmee niet voor de fiets. De Wegenverkeerswet geldt daarentegen wel weer voor alle deelnemers aan het verkeer en niet specifiek voor de auto. De bepalingen in de Wegenverkeerswet, die specifiek voor auto's gelden, staan hier dus binnen een generieker kader.

Niveau van de kaders

Een derde kwestie waarmee de overheid te maken krijgt bij de regulering van AI betreft het niveau waarop kaders moeten worden vastgesteld. Systeem-technologieën zijn naar hun aard universeel toepasbaar en vereisen derhalve naast nationaal beleid ook internationaal beleid. Illustratief zijn in dit verband natuurlijk de internationale afspraken en verplichtingen over de spanning op en de kwaliteit van elektriciteitsnetwerken, zoals onder meer neergelegd in de UCTE-afspraken. Een recenter voorbeeld vormen de talloze mondiale afspraken die de werking van het internet moeten faciliteren, onder meer over basisprotocollen als TCP/IP, DNS en routingprotocollen. Ook voor AI geldt dat het besef dat veel van de uitdagingen hier mondiaal van aard zijn, partijen ertoe heeft gebracht om in verschillende internationale fora rond de tafel te gaan zitten. In feite zien we dat de proliferatie van nationale AI-richtlijnen gepaard gaat met regulatieve convergentie op internationaal niveau.[661]

Behalve bilaterale initiatieven rond AI, zoals de samenwerking van de EU en Japan, Frankrijk en Canada, en Duitsland en India, zijn er ook verscheidene multilaterale initiatieven om te werken aan gemeenschappelijke regels voor AI. Voorbeelden hiervan zijn de gemeenschappelijke ethische principes voor AI van de OESO, gebaseerd op het concept van 'trustworthy AI', zoals dat ontwikkeld is door de AI HLEG van de Europese Commissie.[662] In juni 2019 kwam ook de G20 met een set van ethische principes, die op hun beurt weer grotendeels gebaseerd waren op die van de OESO.[663]

Ook andere fora houden zich bezig met regels voor AI, zoals UNESCO[664] en de Raad van Europa.[665] Parallel aan dit proces intensiveren verscheidene

660 Zo ook Raad van State 2021: 124.
661 Smuha 2019.
662 Zie: High-Level Expert Group on Artificial Intelligence 2019b.
663 OESO z.d.
664 UNESCO z.d.
665 Raad van Europa z.d.

landen hun inspanningen op het terrein van de internationale standaardisatieorganisaties. Deze organisaties houden zich bezig met de technische aspecten van AI maar proberen in toenemende mate ook ethische aspecten op te nemen in hun werk. Vooral waar het om standaardisatie gaat, zijn er ook andere overwegingen waarom landen de samenwerking op internationaal niveau zoeken. Een land als China heeft expliciet de ambitie om actief deel te nemen aan het proces van mondiale standaardisatie, in het bijzonder aangaande gezichtsherkenning. De directeur van Huawei is bijvoorbeeld voorzitter van het IEC Joint Technical Committee for IT van de ISO, wereldwijd een van de belangrijkste standaardisatieorganen. De VS en de EU hebben soortgelijke ambities en streven ernaar hun visie op AI binnen deze organen te promoten. Dit betekent dat standaardisatieorganen als de ISO, de IEEE en de ITU inmiddels het strijdtoneel zijn geworden van landen die proberen de eigen standaarden als mondiale standaard geaccepteerd te krijgen, om zo de eigen bedrijven een competitief voordeel te geven.[666] Hierover gaat hoofdstuk 8.

De toepasbaarheid van de bestaande dan wel het ontwerpen van nieuwe kaders is dus niet alleen een inhoudelijke vraag, maar hangt ook samen met het niveau waarop AI-gerelateerde vraagstukken spelen en, in samenhang daarmee, de strategische ambities van individuele landen. En ook op dit punt zijn de implicaties van het systeemkarakter van AI zichtbaar. De regulering van autonome wapensystemen zal, gezien de koppeling daarvan met oorlogsvoering, op een ander niveau moeten plaatsvinden dan de regulering van medische hulpmiddelen met AI. Hiervoor heeft de Europese Unie van oudsher eigen toelatingskaders, waarbij in Nederland onder andere de Inspectie Gezondheidszorg en Jeugd betrokken is. Alhoewel internationale positionering cruciaal is gegeven de overgang van AI van lab naar samenleving – we zullen dat ook in het volgende hoofdstuk zien – hangt de uiteindelijke keuze voor het niveau van AI-regulering dus eveneens samen met talloze andere – nationale – kwesties. Voor een aantal kwesties heeft de Europese Commissie inmiddels met de concept-Verordening AI een duidelijk signaal afgegeven wat betreft het niveau waarop deze gereguleerd dienen te worden: namelijk op Europees niveau (zie tekstbox 7.4).

666 Smuha 2021: 21.

Tekstbox 7.4 – Niveau van de kaders en de concept-Verordening AI
Met de keuze voor het instrument van de verordening kiest de Europese Commissie voor zowel een horizontaal kader als de directe toepassing wat betreft de regulering van AI, risicovolle AI-toepassingen in het bijzonder. De grondslag van de verordening ligt daarbij in het Verdrag betreffende de werking van de Europese Unie, dat is bedoeld om de interne markt te versterken. Een belangrijk argument van de Commissie voor deze keuze is dat ze een goed functionerende interne markt voor AI-systemen wil faciliteren om zo marktversnippering te voorkomen.[667] Door bovendien de waarden en fundamentele rechten, zoals erkend door de EU, te borgen beoogt de verordening zowel burgers het vertrouwen te geven dat ze AI-toepassingen kunnen omarmen als bedrijven een duidelijk signaal af te geven dat alleen toepassingen op de EU-markt welkom zijn die voldoen aan deze waarden en rechten.[668] Met de grondslag in het Europese interne marktinstrumentarium is echter meteen ook een beperking gegeven, en wel voor AI-toepassingen die een breder maatschappelijk belang dienen. De verordening bevat slechts minimumvereisten om de risico's en problemen in verband met AI aan te pakken en ontbeert daarmee een positieve ethiek. Een deel van het maatschappelijk potentieel van AI, namelijk het deel dat niet of niet zelfstandig via de markt tot stand komt, adresseert de verordening niet.

Ondanks de harmonisatie die de Europese wetgever in gang heeft gezet, zullen bedrijven en burgers geconfronteerd blijven met verschillen in regelgeving tussen landen en daarmee met onzekerheid. Weliswaar behoeft de verordening niet eerst omgezet te worden naar nationaal recht, met alle verschillen tussen landen van dien, maar de realiteit is dat met dit wetgevend kader lang niet alle kwesties geadresseerd zijn. Juist vanwege het systeemkarakter van AI blijft de belangrijke opgave op tafel liggen dat wetgeving in een grensoverschrijdende context niet te grillig is (of wordt) dan wel onvoldoende rechtszekerheid biedt. Juist in een grensoverschrijdende context is rechtsonzekerheid over de op de digitale wereld toepasselijke regels een steeds groter probleem. Niet alleen omdat de regels tussen landen onderling verschillen, maar ook omdat onduidelijk is welke regels – dat wil zeggen de regels van welk land – van toepassing zijn. Dat is bijvoorbeeld het geval wanneer AI-systemen werken met datasets die in cloudapplicaties zijn opgeslagen en er onduidelijkheid bestaat over het recht dat van toepassing is omdat de locatie van de dataset varieert met de beschikbare capaciteit. Maar soms ook verlangen de regels van het ene land een handeling

667 Europese Commissie 2021: 1-4.
668 Floridi 2021.

die door de regels van het andere land wordt verboden. In zijn preadvies voor de Koninklijke Nederlandse Vereniging voor Internationaal Recht waarschuwt de Australische hoogleraar Svantesson voor deze, wat hij noemt, *'hyperregulation'*.[669] Niet verrassend klinkt, ook van anderen dan juristen, de roep om een veel uniformere mondiale wetgevingsagenda. Europa zou daarin het voortouw kunnen nemen, bijvoorbeeld via een European Digital Rule of Law.[670]

Betrokken actoren en hun instrumentarium

Bij de vraag welke kaders nodig zijn voor AI is ten slotte de ontwikkeling van het krachtenveld van groot belang. Waar kan en moet de markt een rol vervullen en is zelfregulering dus aan de orde? Waar kan worden gesteund op de individuele verantwoordelijkheid van burgers en waar en wanneer is daarentegen regulering vanuit de overheid onmisbaar? De erkenning door de overheid dat er ruimte is voor zelfregulering, ligt verankerd in de Aanwijzingen voor de Regelgeving en daarmee kan zelfregulering expliciet een beleidskeuze zijn van de overheid.[671]

Technologiebedrijven ervaren inmiddels de druk en de verantwoordelijkheid om zelf regels op te stellen voor de ontwikkeling en het gebruik van AI. De afgelopen jaren heeft er in dat opzicht een ware omslag plaatsgevonden. Hielden zij regulering voorheen voornamelijk af, nu de kritiek aanzwelt, werken vrijwel alle grote technologiebedrijven aan codes en richtlijnen om duidelijk te maken aan welke regels AI zou moeten voldoen. In sommige gevallen hebben zij daaraan ook concrete voorstellen verbonden, zoals ethische toetsingscommissies. Een steeds groter aantal bedrijven pleit bovendien voor overheidsregulering, onder meer omdat ze lijken te vrezen voor het verlies van marktaandeel als ze stilzitten. Rondom de jaarlijkse bijeenkomst van wereldleiders in Davos in 2020 spraken verscheidene CEO's zich hierover uit. Google's CEO Sundar Pichai gaf aan dat regulering van AI nodig is vanwege de "potentieel negatieve gevolgen". President Brad Smith (en Chief Legal Officer) van Microsoft waarschuwde dat overheden niet moesten wachten totdat de technologie volwassen is om het gebruik ervan te reguleren. Microsoft heeft daarom zelf een commissie ingesteld om met beleidsaanbevelingen te komen. IBM CEO Ginni Rometty kondigde de lancering aan van een intern onderzoekslab om beleidsinitiatieven uit te denken. Google richtte de adviesraad ATEAC op die, echter, vanwege een controverse rond de leden ervan, snel weer opgedoekt werd. Facebook kondigde een interne instantie aan die de media als de Hoge Raad van het bedrijf aanduiden. Deze Oversight Board oordeelde onder meer in mei 2021 over de

669 Svantesson 2020: 121.
670 Hagedoorn 2021: 140.
671 Staatscourant 2017, 69426.

rechtmatigheid van de verbanning van de Amerikaanse oud-president Trump van het platform.

Zelfregulering is een vorm van regulering binnen de eigen organisatie of de structuren waarbinnen partijen opereren. In de meest verregaande vorm betekent zelfregulering dat private actoren zelf de relevante normen en regels opstellen, implementeren, de naleving daarvan controleren en dwang uitoefenen op het moment dat partijen zich daar niet aan houden.[672] Middelen daartoe zijn onder meer private standaarden, vrijwillige programma's, professionele richtlijnen, *codes of conduct*, best practices, publiek-private samenwerkingen en certificeringsprogramma's. En volgens sommigen ook procesbenaderingen, waarbij ethische beginselen geprogrammeerd worden in machines en intern toezicht wordt georganiseerd.

Zelfregulering kan als voordeel hebben dat partijen in het veld zich daaraan binden, omdat principes van onderop door eigen initiatief tot stand zijn gekomen. Ook kan – in principe – met veel preciezere standaarden worden gewerkt, omdat partijen de kennis hebben over wat in de praktijk werkt. Zelfregulering hoeft bovendien niet het eindstation te zijn, maar kan ook dienen als tussenstap op weg naar wetgeving. Mede om die reden werken ook diverse publieke instanties met zachtere regulering. Voorbeelden in ons land zijn de Richtlijnen voor het gebruik van algoritmes door overheden, opgesteld door het ministerie van Justitie en Veiligheid. Het Verenigd Koninkrijk stelde een ethische code voor AI voor om schadelijke effecten van premature wetgeving te vermijden.[673] De adoptie van dergelijke, meer van onderop geformuleerde, initiatieven gaat doorgaans ook veel sneller dan de implementatie van wetgeving en kan daarmee een nuttig instrument zijn om urgente vraagstukken te adresseren. Ook kunnen ze gemakkelijker worden herzien, of worden afgeschaft als ze overbodig blijken.

Potentieel problematisch bij zelfregulering is natuurlijk het democratisch tekort, en daarmee de legitimiteit van de gestelde regels. Een belangrijke tekortkoming van zelfregulering is bovendien dat door private partijen opgestelde regels moeilijker afdwingbaar zijn.[674] Deelname is dan ook vaak onvolledig. Bij AI zijn het vooral welwillende partijen die participeren in initiatieven, om de technologie in lijn te brengen met ethische principes en maatschappelijke waarden. Problematischer en omstreden toepassingen blijven zodoende echter ongereguleerd. Bovendien levert de enorme proliferatie van handvesten,

672 Zie onder meer Giesen 2007; Smits 2015.
673 Select Committee on Artificial Intelligence 2018.
674 Overigens kan de rechter de toepasselijke rechtsregels interpreteren en daar invulling aan geven in het licht van door partijen afgesproken zelfregulering. Zie: Giessen 2007.

richtlijnen en dergelijke een aanzienlijk coördinatieprobleem op. Want welke van deze documenten moet nu precies worden gevolgd, en wat te doen met eventuele spanningen daartussen? Wie is in dat geval scheidsrechter?

Gary Marchant voorspelt dat *'soft law measures'* de komende jaren de default zullen zijn, omdat AI zich razendsnel ontwikkelt en over de wereld verspreidt. De overheid kan hooguit hier en daar wat kleinere problemen oplossen. Het is volgens Marchant daarom nodig om te onderzoeken hoe vormen van zelfregulering voor AI die we nu zien opkomen, indirect kunnen worden afgedwongen en gecoördineerd. Het is een punt dat ook Bert-Jaap Koops maakt in een al wat ouder stuk over ICT-regulering. Koops stelt dat pure zelfregulering in de praktijk nauwelijks voorkomt. Er is vaker wel dan niet ook een rol weggelegd voor de overheid.[675] Zelfregulering en overheidsregulering staan in de praktijk vaak naast elkaar, vullen elkaar op belangrijke punten aan en versterken elkaar. Te denken valt aan samenwerking tussen private actoren en de overheid, bijvoorbeeld om AI-strategieën op te stellen, zoals in veel landen gebeurde.[676] Of aan vormen van zelfregulering waarbij normen wettelijk zijn vastgelegd en verdere details sectoraal worden uitgewerkt. Daarbij moet er volgens Koops een combinatie zijn van consistente druk vanuit de overheid en de samenleving, van beloning voor sociaal wenselijk gedrag, en van manieren om de vinger aan de pols te houden of dat gedrag niet slechts voor de bühne is.

Juist hier lijkt de schoen steeds meer te wringen. Niet specifiek voor AI, maar wel relevant voor ons betoog is de kritiek dat de Hoge Raad bij de toetsing van zelfregulering weinig standvastig en in ieder geval niet leidend lijkt te zijn.[677] Belangrijker wellicht nog wel is dat de wetgever zich nogal afzijdig lijkt te houden en soms zelfs een terugtrekkende beweging maakt.[678] Giesen merkt daarover op: er is "ten minste één goede reden om te verdedigen dat wetgevend terugtrekken niet te verantwoorden is. Namelijk deze dat de wetgever op die manier de regie kwijtraakt. Als je de private partijen zelf aan zet laat, staat de wetgever buitenspel. En dat terwijl het in elk geval tot de verantwoordelijkheid van de wetgever behoort om eindverantwoordelijk voor het complete wetgevingsproces te zijn. Dat begint bij regie voeren."[679]

De bovenstaande constatering is van groot belang voor de ontwikkeling van AI, nu al voor diverse kwesties duidelijk is dat zelfregulering niet volstaat. Zo kampt AI-ontwikkeling in technisch opzicht nog met veel problemen die een

675 Koops et al. 2006.
676 Mols 2019.
677 Menting 2016.
678 Giesen 2018.
679 Giesen 2018: 137.

betrouwbare toepassing ervan in de weg zitten. Zelfregulering veronderstelt dat ontwikkelaars in staat zijn om producten op de markt te brengen die voldoen aan door de sector ontwikkelde standaarden, waaronder betrouwbaarheidseisen. Volgens AI-onderzoekers Gary Marcus en Ernest Davis is de sector daartoe echter nog onvoldoende in staat. Vooral het ontbreken van goede ontwikkelpraktijken is daar volgens hen debet aan. AI-onderzoek richt zich op kortetermijnoplossingen, zoals onmiddellijk werkende code, maar zonder de laag van technische waarborgen die andere velden kenmerkt. Stresstests zijn er bijvoorbeeld nauwelijks en *machine learning*-systemen met adequate risicomarges ontbreken ten ene male.[680] Bovendien voorzien goede ingenieurs hun producten altijd van terugvalopties, zoals dubbele remmen en meerdere besturingssystemen, en opties om ze veilig te laten falen. Ook dergelijke terugvalopties zijn zeldzaam bij AI.

Deze aanpak, gericht op het op grote schaal op de markt brengen van nieuwe, maar zelden volledig uitontwikkelde producten is een belangrijke verklaring voor het mondiale succes van de grote technologiebedrijven. Gebruikers zorgen er daarbij voor dat producten geoptimaliseerd en verder ontwikkeld worden. Dit ontwikkelingsmodel staat diametraal tegenover dat van bijvoorbeeld auto's, medicijnen en vliegtuigen, die eerst uitgebreid worden getest alvorens ze in gebruik worden genomen. Natuurlijk: software kan in tegenstelling tot veel fysieke producten gemakkelijk vanaf een afstand worden gewijzigd en geüpdated en is bij mankementen derhalve niet onderworpen aan kostbare terugroepacties. Deze praktijk komt echter met een toenemend aantal risico's wanneer toepassingen steeds meer verweven raken met processen in de fysieke wereld, en – zoals bij AI – ten grondslag komen te liggen aan beslissingen die een grote impact hebben op mensenlevens. AI zal daarom, zoals ook de Europese Commissie bepleit, in een steeds groter aantal situaties moeten gaan voldoen aan een aantal basale betrouwbaarheidsvereisten om überhaupt in gebruik te mogen worden genomen.

Om zelfregulering een effectieve en legitieme optie te doen zijn, zal AI hiernaast moeten voldoen aan een bredere set van beginselen die de samenleving belangrijk vindt en die zijn gecodificeerd in nationale en internationale verdragen. De veelal in AI-richtlijnen geformuleerde principes trechteren naar een set van principes die veel overeenkomsten vertonen met de principes die gangbaar zijn in de medische ethiek. Namelijk respect voor de menselijke autonomie, het

680 De auteurs gebruiken het voorbeeld van een lift die altijd een veel groter gewicht kan dragen dan berekeningen suggereren en servers die meer internetverkeer aankunnen dan dagelijks nodig is.

voorkomen van schade, eerlijkheid en uitlegbaarheid.[681] Deze principes vormen de kern van onder meer de OESO-principes voor AI en de High Level Group on Artificial Intelligence van de Europese Commissie. De veelgemaakte verwijzing naar de medische ethiek kan echter niet verhelen dat aan veel voorwaarden voor de implementatie van die principes momenteel niet is voldaan.[682]

AI-ontwikkeling kent, in tegenstelling tot de medische wetenschap, om te beginnen geen gemeenschappelijk doel als het bevorderen van de gezondheid en het welzijn van de patiënt. Zoals we in deel 1 hebben gezien, is AI-ontwikkeling, in tegenstellig tot de medische professie, bovendien een jonge professie, met een zeer korte professionele geschiedenis en daardoor nog nauwelijks helder gearticuleerde normen van goed gedrag. AI-ontwikkelaars komen, in tegenstelling tot medische beroepsbeoefenaars, bovendien uit verschillende disciplines en professionele achtergronden. Deze kennen incongruente geschiedenissen, culturen, prikkelstructuren en morele verplichtingen. Softwareontwikkeling, de meest aanverwante discipline, is geen wettelijk erkend beroep met verplichtingen tegenover de samenleving, onder meer omdat een systeem van licenties en helder gedefinieerde professionele zorgplichten ontbreken. De twee grootste beroepsorganisaties, het Institute of Electrical and Electronic Engineers (IEEE) en de Association of Computing Machinery (ACM), hebben weliswaar verschillende codes gepubliceerd en deze herhaaldelijk herzien, maar deze codes zijn relatief kort, theoretisch en bevatten geen adviezen of specifieke gedragsnormen.[683]

Misschien nog wel het belangrijkst is dat AI-ontwikkeling niet gebonden is aan AI-specifieke juridische en beroepsmatige verantwoordingsmechanismen. Verhaal en herstel zijn momenteel niet of nauwelijks mogelijk. Enkele uitzonderingen zijn privacyschendingen en datalekken, maar het kenmerk daarvan is nu juist dat verhaal en herstel hier door wetgeving (AVG) worden gereguleerd. Dat laat het behartigen van de overige waarden over aan de zelfregulering van private partijen, waarvan het langetermijncommitment aan die waarden geenszins gegeven is.[684] Dit wringt des te meer nu de discussie over AI, zoals die gevoerd wordt door het bedrijfsleven en in sommige delen van de wetenschap, zich hoofdzakelijk richt op de vraag hoe en onder welke voorwaarden AI te gebruiken. De wenselijkheid van dit gebruik als zodanig is echter niet of nauwelijks onderwerp van gesprek.[685]

---

681 Floridi en Cowls (2019) betogen dat de huidige AI-principes het meest overeenkomen met principes uit de bio-ethiek en voegen daar zelf uitlegbaarheid als nieuw principe aan toe.
682 Mittelstadt 2019: 503.
683 Mittelstadt 2019: 503.
684 Voor verwijzingen zie: Mittelstadt 2019: 504.
685 Greene et al. 2019.

Kenmerkend voor een systeemtechnologie als AI is bovendien dat de introductie daarvan in de samenleving in kwesties resulteert die het domein van de technologiebedrijven en andere private partijen overstijgen. Technologiebedrijven stellen hoofdzakelijk technische oplossingen voor om problemen met AI te verhelpen.[686] Dat is logisch, want daar ligt ook hun expertise. Maar de reikwijdte van die oplossingen is te beperkt voor een betekenisvolle aanpak van die problemen. Discriminatie, bijvoorbeeld, is in de eerste plaats een maatschappelijk probleem. Dat probleem vraagt om oplossingen op het terrein van onder meer de toegang tot instituties en een normatief debat over welke vormen van onderscheid we maatschappelijk aanvaardbaar vinden. Ten tweede zijn er vraagstukken die niet op het niveau van individuele bedrijven en toepassingen spelen dan wel niet op voorhand adequaat door hen te adresseren zijn. Zelfs wanneer bedrijven zich aan alle relevante wet- en regelgeving houden, kunnen dergelijke tweede- en derde-orde-effecten van AI optreden. Denk aan de verandering van de werkgelegenheid en de daarmee samenhangende noodzaak van scholing. Ten derde gaat het om een discussie over de doelen waarvoor we AI-toepassingen wel en niet willen gebruiken. Denk bijvoorbeeld aan de invoering van autonome wapensystemen. Die discussie kunnen technologiebedrijven niet zelf voeren of op z'n minst niet alleen voeren, omdat de (gewenste) uitkomst daarvan mogelijk indruist tegen het eigen bedrijfsbelang. Ten vierde is zelfregulering geen optie als mensenrechten en fundamentele normen en waarden van de democratische rechtstaat op het spel staan.[687] En daarvan is volgens verschillende onderzoeken inderdaad sprake.[688]

Wat betreft de normering van AI door de overheid, zal de discussie dus niet alleen moeten gaan over de kenmerken van de technologie zelf (is deze betrouwbaar, veilig en voldoende transparant en uitlegbaar?) en het handelen van bedrijven en organisaties die AI ontwikkelen en toepassen. Een dergelijke discussie is te beperkt nu het bij AI om een systeemtechnologie gaat. Evenzeer zal het moeten gaan over de doelen die we als samenleving willen nastreven, en daarmee over de vraag waar, waarvoor en onder welke condities we AI willen gebruiken.[689] Daarbij hoort ook het beperken en zelfs verbieden van het gebruik van AI op bepaalde terreinen, waar de Europese concept-Verordening AI inmiddels een voorzet toe geeft.

686 Häußermann en Lütge 2021. Zie ook: Hagendorff 2020. Hagendorf constateert tevens dat hoe groter het aantal mannen is dat betrokken is bij het opstellen van ethische richtlijnen, des te vaker technische oplossingen naar voren worden gebracht.
687 Vetzo et al. 2018.
688 European Union Agency for Fundamental Rights 2020; Hirsch Ballin 2021.
689 Floridi et al. 2018.

Het systeemkarakter van AI resulteert bovendien in overlap tussen sociale, politieke, commerciële en onderzoeksbelangen. Een enkele partij of groep van partijen kan deze belangen niet tegelijkertijd behartigen. Dit betekent dat het voor een individuele actor onmogelijk maar ook ongewenst is om een monopolie te hebben op de ethiek van AI en de agenda te domineren als het gaat om de kaders waaraan AI moet voldoen. Om te voorkomen dat de private sector en deels ook de wetenschap de standaard stellen ten aanzien van wat een goede AI-samenleving is, is volgens auteurs als Corinne Cath een meer 'gedurfde' strategie noodzakelijk. In zo'n strategie wordt het volledige spectrum van unieke uitdagingen van AI voor de samenleving in termen van eerlijkheid, sociale gelijkheid en verantwoordingen geadresseerd.[690] Bij het formuleren van die strategie zouden, zoals ook in hoofdstuk 6 is betoogd, alle partijen betrokken moeten zijn die door AI worden geraakt.[691] Met deze laatste constatering komt als vanzelf ook de overheid in beeld, die immers als taak heeft integrale afwegingen te maken en daarbij de belangen van verschillende partijen in ogenschouw te nemen. Ook de Europese concept-Verordening AI toont dat na een dergelijke integrale afweging nog stappen vereist kunnen zijn om de inbreng van andere partijen te waarborgen, onder meer bij de concretisering van wettelijke vereisten en het aankaarten van onrecht en schade (zie tekstbox 7.5).

Tekstbox 7.5 – Actoren en de concept-Verordening AI

De Europese Commissie heeft met het voorstel voor de AI-Verordening het initiatief tot regulering overduidelijk naar zich toe heeft getrokken, met als consequentie dat de ontwikkelingen in de markt zich hiernaar hebben te richten. Toch is de rol van private partijen daarmee zeker (nog) niet uitgespeeld. Een onderbelicht aspect binnen de verordening is dat een groot deel van de regulering van AI, in het bijzonder rond hoogrisicosystemen, komt te liggen bij standaardisatieorganisaties als de CEN (Comité Européen de Normalisation) en de CENELEC (European Committee for Electrotechnical Standardisation.[692] De concept-Verordening AI vereist namelijk dat partijen die AI-systemen op de markt willen brengen, nader vast te stellen AI-standaarden raadplegen. Deze standaarden betreffen onder meer de vereiste om een kwaliteitssysteem in te richten, technische documentatie aan te leggen, menselijk toezicht te organiseren en in 'logging' te voorzien.

690 Cath et al. 2018.
691 Cath et al. 2018: 523.
692 Veale en Zuiderveen Borgesius (2021) betogen dat er in de praktijk nauwelijks situaties zullen zijn waarin gebruik wordt gemaakt van de in de verordening herhaaldelijk genoemde onafhankelijke *notified bodies*, geaccrediteerd door nationale toezichthouders. Op het moment dat er standaarden zijn, is namelijk alleen nog een zelf-assessment vereist.

Maar het proces van standaardisatie is gevoelig voor lobby vanuit het bedrijfsleven. Bovendien hebben belangenorganisaties veelal niet de middelen en de expertise om hierin te participeren. Zo is er kritiek dat het raamwerk (het zogenoemde New Legislative Framework) dat ten grondslag ligt aan de concept-Verordening, de belangen van consumenten onvoldoende dient. Een belangrijke kwestie is voorts dat hoogrisicotoepassingen raken aan een groot aantal fundamentele rechten, een terrein waarop standaardisatieorganen weinig kennis en ervaring hebben.

Genoemd moet ook worden dat in het concept van de AI-Verordening (nieuwe) procedurele rechten voor individuen en belangenorganisaties ontbreken, bijvoorbeeld om een klacht in te dienen, compensatie te zoeken of uitkomsten te betwisten.[693] Op andere terreinen blijken dergelijke instrumenten een belangrijke aanjager voor de ontwikkeling van jurisprudentie, zeker wanneer invloedrijke bedrijven de beleidsvorming teveel lijken te domineren en tegenwicht noodzakelijk is. In het huidige voorstel kunnen alleen de bedrijven die aan de vereisten van de verordening zijn onderworpen, de besluiten van de overheid aanvechten. Aangezien er bij AI fundamentele rechten in het spel zijn, roept dat de nodige vragen op over de mate waarin ook andere dan deze bedrijven invloed zouden moeten kunnen uitoefenen.[694]

De Nederlandse overheid benadrukt in haar reactie op de verordening de noodzaak dat voor "burgers en consumenten duidelijk is hoe zij hun recht kunnen halen' en ziet graag dat dit 'in specifieke (consumenten) regelgeving alsnog gebeurt".[695]

Dat de overheid een verantwoordelijkheid heeft bij de regulering van de inbedding van AI in de samenleving, blijkt uit het eerste deel van dit hoofdstuk. Zij heeft zich daarbij te oriënteren op het instrumentarium dat daartoe wordt ingezet, waaronder de kenmerken en het niveau van regulering en de mate waarin private partijen bereid en in staat zijn om zelf publieke waarden te behartigen.

Dat de overheid een rol heeft, zegt nog weinig over de reikwijdte en intensiteit van die rol. Daarover gaat het tweede deel van dit hoofdstuk. Gebaseerd op de

693 European Data Protection Board en European Data Protection Supervisor 2021.
694 Vergelijk Smuha 2021.
695 Dit standpunt is te vinden in Fiche 2: Verordening betreffende Kunstmatige Intelligentie, van het Ministerie van Economische Zaken en Klimaat, Ministerie van Justitie en Veiligheid en Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

geschiedenis van eerdere systeemtechnologieën, betogen we hierin dat het proces van inbedding daarvan gepaard gaat met een steeds grotere rol van de overheid. Die rol blijft daarbij niet beperkt tot de regulering van AI zelf en de acute problemen die daaromheen spelen, maar betreft vooral ook de co-evolutie van technologie en samenleving op de lange termijn en de structurele uitdagingen, kansen en risico's die daarmee samenhangen.[696] In feite hebben we het dan over overheidsregulering als instrument bij de inrichting van onze leefomgeving, namelijk de 'digitale leefomgeving', en de wisselwerking tussen deze omgeving en tal van kwesties die spelen in de fysieke wereld. Alleen vanuit een dergelijk breder en toekomstgericht perspectief op regulering zal de overheid voldoende in staat zijn om blijvend verantwoordelijkheid te dragen voor de behartiging van publieke waarden.

Kernpunten – Normering van AI door de overheid

- Het systeemkarakter van AI brengt een variëteit aan sociale, politieke, commerciële en onderzoeksbelangen samen. Integrale afwegingen, garanties voor publieke waarden en oog voor belangen van verschillende partijen kunnen niet zonder een sturende rol van de overheid.
- Als systeemtechnologie zal AI alomtegenwoordig worden. Dat vraagt van de overheid dat ze in staat is het volledige spectrum van uitdagingen van AI voor de samenleving te overzien en, waar nodig, tijdig met wetgeving te adresseren. Daarbij dient het niet alleen te gaan over de technologie zelf en het handelen van gebruikers, maar ook over de doelen en belangen die de samenleving wil nastreven, en daarmee over de vraag waar, waarvoor en onder welke condities we AI willen gebruiken.
- Overheidsregulering van AI kent geen standaardaanpak. Het instrumentarium waarin normen worden neergelegd en het niveau waarop de regels worden afgekondigd (internationaal, nationaal, decentraal), zullen telkens een zoektocht zijn naar een goede balans tussen enerzijds een effectieve waarborg van publieke waarden en anderzijds voldoende ruimte voor innovatie.
- Wil de overheid gegeven deze uitdagingen effectief, tijdig en betekenisvol kunnen interveniëren met aandacht voor eenheid van beleid, dan is een bredere wetgevingsstrategie onontbeerlijk.

696 Zie bijvoorbeeld: Just en Latzer 2017; Krupiy 2020.

7.2 AI-regulering en de digitale leefomgeving

Bij de regulering van eerdere systeemtechnologieën legde de overheid gaandeweg een steeds groter gewicht in de schaal. Nieuwe technologie krijgt aanvankelijk vaak de ruimte om zich te ontwikkelen. Maar naarmate die technologie breder verspreid raakt in de samenleving, daarin dieper ingrijpt en de impact ervan meer aan het licht komt, is het vaak nodig verdere eisen te stellen. Een voorbeeld is het moment waarop het toenemende gebruik van auto's tot gevaarlijke toestanden en de eerste verkeersslachtoffers leidde, en tot grootschalige protesten in zowel de VS als Europa. In de VS won daarbij de autolobby de strijd en werd de auto het dominante vervoermiddel, terwijl in Europa ook openbaar vervoer tot stand kwam en veel explicieter ruimte werd gecreëerd voor voetgangers, om zo ook financieel minder draagkrachtige groepen in hun transport en mobiliteit te faciliteren.

De effecten, kansen en risico's rond een systeemtechnologie veranderen dus gaandeweg van aard. De aandacht van de regulering verschuift daardoor van de technologie zelf naar tevens de regulering van bredere effecten, zoals andere economische dynamieken en de context waarin de technologie wordt toegepast. Het ging dus niet meer alleen om de eisen aan de verbrandingsmotor of de stroomsterkte, maar ook om verkeersregels voor gemotoriseerd verkeer, overlast in de binnensteden, veilige consumentenelektronica en de verduurzaming van energiecentrales toen bleek dat de verbranding van fossiele brandstoffen een belangrijk effect heeft op de opwarming van de aarde. Dergelijke voorbeelden tonen bovendien dat telkens verschillende afwegingen mogelijk zijn, waarbij bedrijven en belangengroepen het inbeddingsproces naar hun hand proberen te zetten.

Hoewel de overheid gaandeweg een groter gewicht in de schaal legt, is niet bij voorbaat duidelijk hoe groot dat gewicht bij AI dient te zijn. Met andere woorden: welke omvang en intensiteit de regulering van AI dient te krijgen. Bij de inbedding in de samenleving van een systeemtechnologie gaat het om een proces dat decennia in beslag neemt. Dat proces brengt veel onzekerheid met zich mee, vooral over de maatschappelijke impact die de technologie zal gaan hebben en hoe regulering die impact in banen kan leiden. Hieronder staan we allereerst stil bij die onzekerheid. Deze kan aanleiding zijn om zowel te wachten met maatregelen als vroegtijdig in te grijpen om eventuele problemen te voorkomen. De goede 'timing' van overheidsinterventie is daarom een tweede thema dat we bespreken. In dat verband staan we tevens stil bij de noodzaak dat de overheid alert is op sturende krachten, vooral in de markt. Gaandeweg de inbedding van een technologie wordt het voor de overheid namelijk steeds moeilijker om de regulerende invloed van andere partijen nog te keren of in een andere richting te sturen.

De laatste kwestie die de omvang en intensiteit van overheidsregulering sterk beïnvloedt, is de wisselwerking tussen een systeemtechnologie, in dit WRR-rapport dus AI, en bredere ontwikkelingen en uitdagingen in de samenleving. Een systeemtechnologie doet iets met een samenleving en een samenleving doet iets met een systeemtechnologie. Wat precies dat 'iets' is en of dat 'iets' in verband wordt gebracht met ontwikkelingen die ogenschijnlijk weinig met AI van doen hebben (zoals de klimaatopgave of de houdbaarheid van de zorg), hangt in belangrijke mate af van hoe de overheid zich opstelt. Vanuit drie besproken kwesties komen we tot de conclusie dat het dringend nodig is niet alleen zaken als privacy, aansprakelijkheid, transparantie, verzekerbaarheid, consumentenbescherming te reguleren maar ook de digitale leefomgeving zodanig in te richten dat AI daarbinnen ook op de langere termijn ten goede kan komen aan de publieke waarden.

Onzekerheid

De omgang met onder andere de verbrandingsmotor en elektriciteit toont dat de kaders daarvoor niet van de ene op de andere dag tot stand komen en bovendien in de loop van decennia vaak weer kunnen veranderen. Illustratief voor dit zoekende proces zijn bij AI de Richtlijnen voor het gebruik van algoritmes door overheden, opgesteld door het ministerie van Justitie en Veiligheid. Als richtlijnen worden daarin genoemd: bewustzijn van risico's, uitlegbaarheid, gegevensherkenning, auditeerbaarheid, verantwoording, validatie en toetsbaarheid. Onderzoekers van Waag deden onderzoek naar de toepasbaarheid van deze richtlijnen in de praktijk.[697] Hun voornaamste conclusie was dat ze scherper geformuleerd zouden moeten worden om van betekenis te kunnen zijn. De richtlijnen maken bijvoorbeeld niet duidelijk hoe verantwoordelijkheid vorm krijgt en voor welke partijen algoritmes uitlegbaar moeten zijn. Ook worden ze niet gekoppeld aan de richtlijnen en standaarden in de reeds aanwezige wet- en regelgeving, zoals de Algemene beginselen van behoorlijkheid van bestuur.

De stappen die het ministerie van Justitie en Veiligheid heeft gezet en de verdere uitwerking daarvan zoals gesuggereerd door Waag zijn van groot belang wanneer een systeemtechnologie zojuist de overgang van het lab naar de samenleving heeft gemaakt. Op dat moment spelen vaak heel specifieke en concrete inbeddingsvragen. Voorbeelden zijn die naar de aansprakelijkheid voor schade, verzekerbaarheid, rechtshandelingen door autonoom handelende systemen en auteursrecht op algoritmes. In veel gevallen zijn die vragen te beantwoorden door op de bestaande kaders terug te grijpen en deze vervolgens te actualiseren. Het is in deze fase in veel opzichten nog te vroeg voor speciale regels voor AI. Ook de oprichting van zoiets als een AI-autoriteit

697 Waag 2020.

of een specifieke AI-toezichthouder is op zo'n moment voorbarig, omdat een helder afgebakend werkterrein ontbreekt. Er is dan nog onvoldoende bekend over de generieke patronen bij bijvoorbeeld risico's bij de inzet van AI, de noodzakelijke rechtsbescherming of mededinging en marktordening.

Bij een systeemtechnologie als AI zal de overheid regulering aanvankelijk dus veelal incrementeel benaderen. Voor bekende risico's worden regels al redelijk snel na de introductie van een systeemtechnologie verhelderd of aangepast, of er worden nieuwe regels geïmplementeerd. Dit zien we bij AI op verschillende plekken gebeuren, alhoewel het tempo niet erg hoog ligt en aanpassing niet systematisch gebeurt. Maar zowel de ingewikkelde technologische structuur van AI als de koppeling met specifieke gebruikscontexten maken de technologie deels onkenbaar en brengen een grote mate van complexiteit en daarmee onbekende risico's met zich mee.[698] Deze onbekende risico's zullen pas bij intensiever en grootschaliger gebruik aan de oppervlakte komen en vereisen zorgvuldige monitoring, bijvoorbeeld door signalementen en foutregisters.[699]

Vooral partijen die dicht op de ontwikkelingen zitten, krijgen vaak als eerste zicht op de vragen en problemen waarmee de inbedding van AI gepaard gaat. Naast de in hoofdstuk 6 besproken maatschappelijke organisaties vervullen ook rechters, toezichthoudende instanties en het parlement deze signalerende rol.[700] Parlementsleden kunnen als vertegenwoordigers van de samenleving geluiden opvangen over incidenten die mogelijk wijzen op knelpunten met publieke waarden. Rechters krijgen zaken aangedragen waarin partijen als gemeenten van algoritmes gebruik maken, zoals duidelijk werd in tekstbox 7.1. Inspecties en toezichthouders zien nieuwe toepassingen die op de markt komen, zoals bestuurdersondersteunende systemen in auto's en medische hulpmiddelen op basis van AI, en moeten toezien op processen waarbij steeds vaker AI wordt gebruikt, zoals risicobeoordelingen, cybersecurity, de toekenning van toeslagen of de controle van goederenstromen.

Bovendien hebben dergelijke instanties een maatschappelijke functie bij het signaleren van ontwikkelingen die van invloed zijn op de te borgen publieke belangen en de verhoudingen in het krachtenveld.[701] Voorbeelden hiervan zijn de verkenning die de Autoriteit Financiële Markten en De Nederlandsche Bank verrichtten over AI-gebruik in de verzekeringssector[702] en het recente onderzoek van de Algemene Rekenkamer naar hoe de rijksoverheid algoritmes

698 Burrell 2016.
699 Voor de omgang met onbekende risico's en het voorzorgprincipe, zie: WRR 2008.
700 Vgl. deel IV in Bennett Moses 2007b.
701 WRR 2013.
702 AFM en DNB 2019.

gebruikt.[703] De Rekenkamer constateerde dat een verantwoorde ontwikkeling naar complexe geautomatiseerde toepassingen beter overzicht en een betere kwaliteitscontrole vereist, en ontwikkelde daarom een toetsingskader. Een aantal rijksinspecties en markttoezichthouders heeft daarenboven het initiatief genomen voor een interdepartementale werkgroep om kennis en ervaring met betrekking tot 'toezicht op AI en algoritmes' te delen. Een goede terugkoppeling van de bevindingen van dergelijke instanties naar politiek en bestuur groeit naarmate problemen binnen de bestaande kaders niet zijn op te lossen en/of een domeinoverstijgend karakter hebben, en derhalve generieke maatregelen vereisen.

Timing van overheidsinterventies

Duidelijk is dus dat de overheid, en meer specifiek de wetgever, in eerste instantie zal aftasten en pas in een latere fase een grotere rol zal innemen. Echter, de mogelijkheid om ook daadwerkelijk effectieve eisen te stellen aan de omgang met een technologie, verandert wel mettertijd. Als een technologie als het ware eenmaal vaste voet aan de grond heeft gekregen, is bijsturen ingewikkeld en soms zelfs onmogelijk of ondoenbaar. Dat heeft te maken met wat het Collingridge-dilemma is gaan heten, en met andere partijen dan de overheid die op de inbedding van de technologie sturen en daarmee in feite dus ook een regulerende functie vervullen.

Het Collingridge-dillemma verwijst naar een informatieprobleem en een machtsprobleem bij de overheid. *"When change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult, and time-consuming,"* zo stelde David Collingridge in 1980 in zijn boek *The Social Control of Technology*. In het beginstadium waarin de overheid goed in staat is om de technologische ontwikkeling bij te sturen, zijn de effecten van de technologie nog onvoldoende bekend en is er een groot risico van niet-passende, ineffectieve of zelfs contraproductieve wetgeving. Maar op het moment dat duidelijk is wat er moet gebeuren, omdat die effecten overduidelijk aan het daglicht treden, is de technologie dermate ingeburgerd dat verandering middels sturende wetgeving alleen tegen zeer hoge kosten te realiseren is.[704]

In de loop der jaren heeft het Collingridge-dillemma veel aandacht gekregen. We kunnen hierin een argument lezen voor de stelling dat de overheid bij een nieuwe technologie beter afzijdig kan blijven, in ieder geval in de beginfase. Wanneer een technologie nog in de kinderschoenen staat, is deze kwetsbaar en moet zij daarom vooral omzichtig benaderd en gekoesterd worden. Bovendien

703 Algemene Rekenkamer 2021.
704 Vergelijk Bijlsma et al. 2016.

is de introductie op dat moment nog met veel onbekendheden en onzekerheden omgeven. Maar tegelijkertijd impliceert het Collingridge-dilemma dat het op een later moment te laat is om nog iets te kunnen doen omdat de technologie dan alomtegenwoordig is geworden. Die twee punten hangen logisch samen: wie te laat start, heeft bij de finish veel in te halen – als dat al lukt. In het Collingridge-dilemma schuilt een zekere waarheid, maar die waarheid is tegelijk ook simplistisch, reden waarom Wendel Wallach het een dogma noemt.[705] Het gaat immers voorbij aan de vele krachten die feitelijk inwerken op hoe een technologie gebruikt gaat worden, zowel in vroege als latere stadia van een inbedding in de samenleving.

De sturende werking van technologie

Over deze krachten schreef Lawrence Lessig met het boek *Code* in 1999 een klassieker.[706] Hij betoogde daarin dat digitale technologie behalve door wetgeving, marktmacht en sociale normen ook sterk werd beïnvloed door de technische vormgeving ervan – door programmeercode kortom.[707] Eerder al had Langdon Winner in zijn werk over de politiek van technologie laten zien dat de werking van technologie ook een vorm van regulering is.[708] Dit geldt evenzeer voor AI. Illustratief is de algoritmische moderatie waarmee platformen proactief toezicht houden op de online-inhoud die internetgebruikers met elkaar delen.[709]

Ook de ontwikkeling en toepassing van AI is onderhevig aan verschillende krachten, variërend van de bestaande rechtsregels, de private partijen die de technologie ontwikkelen, tot denkbeelden in de samenleving over autonomie en de menselijke waardigheid, en dus ook keuzes in het ontwerp van AI-systemen, zoals de voorrang die accuraatheid krijgt boven uitlegbaarheid als het gaat om de uitkomsten van die systemen. Standaardisatie en de mate waarin de private sector daarin de toon weet te zetten, zijn een ander voorbeeld van dit type krachten. Toch moeten we bij AI alert zijn op een aanvullende kwestie: de controlerende en daarmee sturende rol en invloed van de mens op de reguleringskracht van de technische vormgeving verandert. Die rol en invloed worden immers problematisch nu AI-systemen zich veelal laten kenmerken door ondoorzichtigheid, complexiteit en zelflerend vermogen.[710]

---

705 Wallach 2015: 71-72.
706 Lessig 2006.
707 Lessig 2006.
708 Winner 1983: 97-111.
709 Over dit algoritmisch toezicht en het Europese beleid in deze, zie Kulk 2020: 132-140.
710 Yeung en Lodge 2019.

Dat sturing in de tijd lastig is, wordt niet veroorzaakt door een deterministische – 'natuurlijke' of 'onvermijdelijke' – ontwikkeling van de technologie zelf, maar door padafhankelijkheid. Padafhankelijkheid laat zich het beste illustreren aan de hand van de werking van het wegennet. Bij de aanleg van nieuwe wegen worden vaak de bestaande routes gevolgd. Die routes zijn echter lang niet altijd de meest efficiënte verbinding van A naar B. Dat die routes desondanks gehandhaafd blijven, komt doordat de bebouwde omgeving zich ernaar heeft gevoegd. Het zou bijvoorbeeld enorm kostbaar zijn om alle woningen en bedrijven langs een weg te verplaatsen. De efficiëntie van een snellere verbinding moet op zo'n moment concurreren met allerlei belangen die voortvloeien uit de keuzes die in het verleden zijn gemaakt, en bij elkaar opgeteld veel gewicht in de schaal leggen. Wie er oog voor heeft, ziet dit proces voortdurend om zich heen gebeuren. De veronderstelde exponentiele groei van de rekenkracht van computers (de wet van Moore) bijvoorbeeld, is geen wetmatigheid maar een zichzelf vervullende voorspelling. Het is niets anders dan een jaarlijks gesteld doel voor ingenieurs, labs en bedrijven, dat vervolgens dicteert dat de digitale infrastructuur moet groeien, dat extra personeel nodig maakt, om de financiering van onderzoek vraagt en de vraag naar snelle halfgeleiders doet groeien.

In een dergelijk proces is er altijd een moment waarop een bepaalde interpretatie van het ontwerp of een gebruik van een technologie de norm wordt en 'closure' plaatsvindt.[711] De controverse daarover verdwijnt en van de verschillende concurrerende ontwerpen blijft er maar een over. Zo reden er rond 1900 verschillende typen voertuigen rond, zowel met elektrische aandrijving als met uiteenlopende typen verbrandingsmotoren.[712] In de loop van de twintigste eeuw bleef daar de op benzine werkende verbrandingsmotor van over, mede vanwege de beperkte actieradius van elektrische vervoermiddelen, een probleem dat ons ook nu bekend in de oren klinkt. Een ander bekend voorbeeld is de fiets, die tegenwoordig twee wielen van dezelfde grootte heeft, aangezien dit de effectiefste manier van voortbewegen oplevert.[713] Aanvankelijk de fiets was een mannelijk en atletisch vehikel, dat veel kracht vereiste. De daarbij horende versie met een zeer groot voorwiel kreeg concurrentie van het huidige model, gericht op veilig en efficiënt transport. De laatste werd na decennia uiteindelijk de norm, waarna de andere varianten van het toneel verdwenen. Overigens geeft de kwaliteit van de technologie lang niet altijd de doorslag bij closure. Illustratief is het eindresultaat van de concurrentie in de jaren zeventig en tachtig tussen de incompatibele standaarden voor videorecorders, VHS, V2000 en

711 Bernstein 2006.
712 Bakker en Korsten 2021: 37.
713 Bijker 1995.

Betamax. Marketingaspecten, zoals prijs en ondersteuning, en niet de kwaliteit van het systeem bleken doorslaggevend voor de uiteindelijke keuze voor VHS.

Closure heeft verregaande gevolgen: alternatieven verdwijnen en het debat daarover komt stil te liggen. Mensen, relaties, andere technologieën en de bestaande praktijken en procedures: alles komt vanaf het moment van closure in een nieuwe verhouding tot elkaar te staan. Tegelijkertijd is uit dit proces de les te trekken dat er tussen de introductie en de verankering van een technologie in de samenleving altijd een of meerdere kantelpunten zijn. Op deze kantelpunten worden de problemen rond een technologie zichtbaar, nog voor het te laat is om er wat aan te doen.[714] Technologische verandering komt immers niet uit de lucht vallen, hoe snel de verandering ook gaat. De tijd om er wat aan te doen kan heel kort zijn, maar kan ook jaren duren. Als de technologische ontwikkeling een hoge vlucht neemt, verkleint dat de ruimte om nog te kunnen interveniëren. Maar dit kan ook gebeuren doordat het gebruik van de technologie enorm toeneemt, zoals gebeurd is bij de mobiele telefoon. Juist vanwege het intensieve gebruik is deze een belangrijk platform geworden voor allerlei andere producten en diensten, zoals betalen of het bedienen van de thuisthermosstaat. Er zijn dus verschillende factoren die closure kunnen bewerkstelligen.

Uit het voorgaande kunnen we de conclusie trekken dat wanneer de overheid zich te lang laat gijzelen door de onzekerheidsparadox[715] waarmee iedere nieuwe systeemtechnologie gepaard gaat, zij het risico loopt niet meer in staat te zijn om effectief en betekenisvol te interveniëren. De regulering van het internet, of eigenlijk het grotendeels ontbreken daarvan, kan daarbij als een belangrijke waarschuwing dienen.[716] Een dergelijke situatie heeft vanzelfsprekend grote gevolgen voor een adequate borging van publieke waarden. Immers, andere krachten krijgen volop ruimte en drukken een grotere stempel op de inbedding van de technologie in de samenleving, zoals een handvol Amerikaanse techbedrijven die drukt op de ontwikkeling van socialemediaplatformen.[717] Maar ook krijgen de specifieke krachten van de AI-technologie zelf ruimte, waarmee vervolgens de ruimte voor doorzichtigheid, uitlegbaarheid en daarmee interventie door de mens onder druk kan komen te staan. Cruciaal punt is hier – en daar komt Collingridge weer om de hoek kijken – dat het gaandeweg steeds moeilijker wordt om de invloed van andere krachten dan de overheid nog te keren en AI in een andere richting te sturen.

714 Wallach 2015: 72; Bennet Moses 2007a: 600.
715 Van Asselt et al. 2010.
716 Stikker 2019; Black en Murray 2019.
717 Helberger et al. 2018.

De stap van de EU om te komen tot een verordening die diverse specifieke AI-kwesties regelt, valt in het licht van het Collingridge-dilemma dan ook toe te juichen. Relevant voor de opgave van regulering zoals wij deze hier bespreken, is dat de Commissie via een risicoclassificatie in vier typen[718] strenge of mindere strenge regels toepast afhankelijk van de functie, het beoogde doel en de modaliteiten van de AI-toepassing. Met andere woorden: de Commissie kijkt niet alleen naar de technologie sec en vanuit een generiek oordeel over het gebruik, maar ook naar de specifieke context. Zo wordt biometrische identificatie in beginsel verboden, omdat deze specifieke risico's inhoudt voor de grondrechten, met name de menselijke waardigheid, de eerbiediging van het privéleven en het familie- en gezinsleven, de bescherming van persoonsgegevens en non-discriminatie. In enkele strikt gedefinieerde, beperkte en gereguleerde gevallen, zoals de gerichte opsporing van vermiste kinderen of daders van ernstige misdrijven, mag biometrische identificatie desondanks worden ingezet. Door voor een risicoclassificatie te opteren moeten de Europese en nationale overheden zich vroegtijdig een beeld vormen van problemen en risico's. Zo wordt de mogelijkheid om de ontwikkeling bij te sturen aanzienlijk vergroot. Een dergelijke aanpak is vergelijkbaar met het signaleren van de problemen die al vroeg zichtbaar werden bij de introductie van de verbrandingsmotor (verkeersongelukken bij de eerste auto's op de weg) of de aanleg van de eerste spoorwegen (eigendomsvraagstukken, achterhaalde aannames wat betreft het recht van overpad).

Met de bovenstaande conclusie arriveren we bij de laatste kwestie van de opgave die in dit hoofdstuk centraal staat: de inhoud van de reguleringsagenda. Wat die inhoud betreft laat het voorgaande betoog zien dat – hoe onzeker het proces en de uitkomst van de inbedding van AI in de samenleving ook moge zijn – overheidsregulering zich de komende jaren niet zal kunnen beperken tot slechts wat onderhoud aan bestaande kaders en het opstellen van principes om vervolgens de verdere concretisering over te laten aan toezicht dan wel rechtspraak. De Europese concept-Verordening illustreert de noodzaak van een bredere wetgevingsagenda. Deze stap van de Europese wetgever past in de tendens dat overheden, waaronder ook de Nederlandse overheid, door een steeds groter aantal partijen worden aangespoord om te komen tot een gedegen technologiebeleid.[719] In dat technologiebeleid dienen publieke waarden te worden verankerd, aldus de inbreng van diverse vertegenwoordigers uit het digitale domein tijdens een

718 Het gaat om toepassingen die als onacceptabel worden aangemerkt, met een minimaal risico, een beperkt risico en een hoog risico.

719 Zie onder meer het Digitale Stembusakkoord, op 12 maart 2021 ondertekend door een brede samenstelling van politieke partijen (streng toezicht op algoritmes is één van de elf punten waar de ondertekenaars zich aan committeren).

gesprek met informateur Hamer in de zomer van 2021.[720] Specifiek voor de inzet van AI gaat het bij dit technologiebeleid ook om fundamentele kwesties over de rol van de menselijke controle alsmede de menselijke maat.[721]

Indachtig toekomstige ontwikkelingen reikt de AI-reguleringsopgave kortom verder dan de optelsom van talloze losstaande kwesties, risico's en ontwikkelingen. Juist bij een systeemtechnologie zijn het niet uitsluitend de technologie en de concrete toepassingen die om wetgeving vragen, maar zeker ook de bredere effecten ervan in de samenleving. Bij een dergelijke technologie is het kortom van groot belang dat de overheid zich niet verliest in een fragmentatie van de grote hoeveelheid individuele juridische vraagstukken die op talloze terreinen spelen, maar zich een beeld vormt van de grotere vragen die spelen rond de inbedding van die technologie. Dit vraagt van de overheid dat zij de reguleringsagenda voor AI vooral ook inricht vanuit de vaststelling dat het bij de inbedding van AI in essentie gaat over de verantwoordelijkheid voor en de inrichting van onze (digitale) leefomgeving.

Een wetgevingsagenda voor de digitale leefomgeving

Dat de wetgever de AI-ontwikkelingen de komende jaren vooral dient te beschouwen vanuit een perspectief op de bredere inrichting van de samenleving, is wederom historisch te illustreren. Zowel de komst van de auto en het vliegtuig als de introductie van elektriciteit dwongen de wetgever keuzes te maken die veel verder reikten dan individuele beleidsdossiers. In het geval van auto's en elektriciteit betrof het een visie op de ruimtelijke ordening en de inrichting van de fysieke leefomgeving. En net als elektriciteit is AI behalve een commodificeerbaar goed waarmee economisch voordeel behaald kan worden, ook een goed waarvan de baten ten goede komen aan grote groepen in de samenleving of zelfs de gehele bevolking. Elektriciteit verlengde de dag, maakte woningen veiliger, steden schoner en veraangenaamde het leven in veel opzichten. Van AI mogen soortgelijke baten worden verwacht. Maar dan zal de overheid er wel zorg voor moeten dragen dat de technologie op plekken terecht komt waar zij een optimale bijdrage kan leveren, en wordt ingezet op een schaal en voor doelen die congruent zijn met die van de Nederlandse samenleving.

720 iBestuur, 31 mei 2021.
721 Voor het domein van de sociale zekerheid zie bijvoorbeeld enkele bijdragen in Hirsch Ballin et al. (red.) 2021.

Evenzeer als voor de auto en het vliegtuig, geldt echter ook voor AI dat de technologie energie nodig heeft en dus brandstof slurpt.[722]

In de breedste zin van het woord kan AI ook sterk vervuilend zijn. Zij "heeft veel te bieden. Evenzeer echter, hebben industriële innovatie, intensieve landbouw en chemische industrie ons veel gebracht maar trekken ze tegelijkertijd een wissel op milieu, klimaat en onze planeet. Een te zware wissel, naar nu blijkt. Zo ook gaan bij digitalisering innovatie, voortuitgang en groei van economie en samenleving hand in hand met risico's. En ook hier betreft het niet alleen consequenties voor individuele burgers en bedrijven, maar evenzeer effecten op collectieve goederen en waarden".[723] Kortom, naarmate de rol van AI in onze samenleving groter wordt, gaan ook steeds vaker fundamentele keuzes over de inrichting van de samenleving alsmede de 'digitale leefomgeving' spelen.

Illustratief voor dergelijke keuzes is het pleidooi van Hirsch Ballin om het beleid voor AI vooral te verbinden aan de bevordering van *resilience* in de levensloop van mensen, als antwoord op hun inherente en onvermijdelijke kwetsbaarheid. In zijn analyse van de verhouding tussen AI en de rechten van de mens formuleert Hirsch Ballin een drietal ijkpunten voor AI-beleid. Twee daarvan zijn illustratief voor de inrichtingsvraagstukken waar wij op doelen. Ten eerste: de keuze om artificiële intelligentie vooral ook te verbinden met levensprojecten van mensen. Kortom, AI-beleid ontwikkelen "om de complexe processen te corrigeren die mensen beletten zich te voeden, te scholen en anderszins hun levensprojecten te ontwikkelen."[724] Ten tweede de noodzaak om zowel het doel waartoe AI wordt ingezet, als de inrichting ervan in alle fases te "verbinden met humaniteit, respect voor verscheidenheid, en steun voor vrij aanvaarde levensprojecten."[725]

Eerder al heeft de overheid op geheel andere terreinen dan AI – ontwikkelingen op terreinen die de samenleving in fundamentele zin raken en ordenen – talloze losse beleidsdossiers samengebracht om deze in relatie tot elkaar te bezien als breder inrichtingsvraagstuk. Ruimtelijke ordening is hiervan een prominent voorbeeld. De afgelopen decennia hebben verschillende kabinetten opvolgende

722 Al langer is bekend dat iedere zoekopdracht op Google te vergelijken is met een flink aantal seconden energieverbruik van een gloeilamp (naar eigen opgave van Google stond een zoekopdracht gelijk aan 17 seconden gebruik van een gloeilamp van 60 watt). Het jaarlijks energieverbruik van talloze landen ligt lager dan dat van het bitcoinnetwerk en talloze windmolens draaien uitsluitend en alleen voor datacenters van bedrijven als Google en Microsoft. Vergelijk Coeckelbergh 2020 en Van Wynsberghe 2021.

723 Prins 2018: 1563.

724 Hirsch Ballin 2021: 33.

725 Hirsch Ballin 2021: 34.

nota's gepresenteerd met daarin verwoord de beleidsopgaven en beleidsdoelen, sturingsfilosofie, instrumentatie en uitvoering.

Illustratief voor een meer op de problematiek van AI en de digitale leefomgeving toegesneden beleidsdocument is de nota *Wetgeving voor elektronische snelweg* uit 1998. Dit 234 pagina's tellende document van het ministerie van Justitie had als doel: "Een legitimatie te geven van het overheidsoptreden tijdens de overgang naar de informatiesamenleving voor zover het instrumentarium van de wetgeving daarbij een rol kan spelen; Die legitimatie te concretiseren in een toetsingskader voor de wetgever; Op belangrijke onderdelen aan te geven wat de verschillen zijn tussen de fysieke wereld en de elektronische omgeving; Voorstellen te doen voor concrete vraagstukken die zich voor doen als gevolg van technologische ontwikkelingen." De huidige Nederlandse Digitaliseringsstrategie (NDS) en de jaarlijkse actualisering daarvan is veel minder expliciet dan beide nota's wat betreft beleidsopgaven, sturingsfilosofie en het instrumentarium. Ook is zij – zeker in tegenstelling tot de nota *Wetgeving voor de elektronische snelweg* – geen agenda die de volle breedte van beleidsterrein 'digitalisering' afdekt. Ondanks de inzet – "een succesvolle digitale transitie in Nederland" – betreft de bulk van de strategie een lijst aan concrete actiepunten, vooral gericht op de publieke sector, geclusterd onder een aantal thema's.

Het risico van dergelijke 'lijstjes' is dat bredere, en veelal veel fundamentelere, ontwikkelingen onvoldoende in beeld komen of niet vanuit een op de lange termijn gerichte visie geadresseerd worden. Hieronder signaleren we drie ontwikkelingen die de WRR cruciaal vindt voor de inbedding van AI in de samenleving. Kernpunt bij deze ontwikkelingen is dat AI in belangrijke mate een 'civiele technologie' is, zoals in hoofdstuk 3 duidelijk werd. AI is een technologie die ontwikkeld is door bedrijven en niet of in ieder geval vaak in mindere mate door de overheid of onafhankelijke wetenschappers. Ook is ze een technologie die gebaseerd is op grote hoeveelheden data, waarvan de overvloedige beschikbaarheid samenhangt met hoe de digitale wereld is vormgegeven, met het internet voorop. Bovendien zijn de partijen die nu al dominant zijn bij het verzamelen, bewerken en verspreiden van data over het internet, tevens de grootste investeerders en ontwikkelaars van AI.[726] Deze factoren drukken een stempel op het inbeddingsproces van AI en de overheid zal zich daartoe moeten verhouden. De wijze waarop de overheid zich de komende jaren hiertoe vanuit haar regulerende taak verhoudt, zal bepalend zijn voor de vraag of publieke waarden voldoende bij die inbedding van AI tot hun recht zullen kunnen komen.

726 Nemitz en Pfeffer 2020.

Drie ontwikkelingen die de inbedding van AI beïnvloeden

Waar gaat het om? Die ontwikkelingen zijn de toename van surveillance in de samenleving, de scheefgroei in het gebruik van digitale middelen en de machtsconcentratie in het digitale domein, met spillover-effecten naar andere terreinen van de samenleving, zoals de samenleving en democratie. Een visie van de overheid op deze ontwikkelingen is van cruciaal belang, omdat ze van grote invloed zijn op de mate waarin AI ook op de langere termijn door de overheid en andere partijen te reguleren valt. Bovendien bepalen deze bredere ontwikkelingen in feite ook hoe de verhoudingen liggen bij de specifiekere kwesties die in verband met de inbedding van AI spelen. Transparantie bijvoorbeeld, is niet alleen belangrijk om inhoudelijk besluiten van individuele AI-systemen te kunnen navolgen, maar ook om te voorkomen dat de ontwikkelaars en gebruikers van die systemen een te grote en tegelijkertijd niet te controleren invloed krijgen op de samenleving.

Surveillance

De eerste ontwikkeling gaat over het grootschalig verwerken van (persoons) gegevens ten behoeve van 'surveillance' en het aan de hand daarvan sturen op het handelen van burgers en bedrijven. Als zodanig is deze ontwikkeling zeker niet nieuw.[727] Dat betekent echter geenszins dat de overheid deze als min of meer als vanzelfsprekend moet zien. De ruimte die bedrijven, organisaties maar ook burgers de komende jaren feitelijk krijgen om tot surveillance over te gaan, zal wezenlijk blijken voor de wijze waarop de digitale samenleving op de langere termijn vorm krijgt. Nu al lijkt observatie, vaak ook heimelijk, steeds meer een vanzelfsprekendheid te zijn, voor bedrijven, overheden, en burgers.[728] Illustratief zijn de deurbellen met gezichtsherkenningstechnologie (zie tekstbox 7.6).

Tekstbox 7.6 – Ring, de slimme deurbel

De digitale deurbel Ring bevat een camera die op de straat is gericht. Het is echter niet toegestaan om permanent de straat te filmen. Een voorwaarde is dan ook dat dit apparaat alleen is toegestaan wanneer de kans op inbraken verhoogd is. Een sluimerend effect van deze waarborg is dat mensen in buurten met veel criminaliteit, vaak armere buurten, in de praktijk minder recht hebben op privacy.

727 Vergelijk de bijdragen in Hildebrandt en Gutwirth 2008.
728 Voor het aanbod, zie bijvoorbeeld bol.com, amazon.com of de mediamarkt via zoektermen als 'afluisterapparaat', 'anti-diefstal', 'beveiligingscamera', 'gezichtsherkenning'.

Een ander voorbeeld van dit soort 'sluimerende groei van surveillance', vinden we in politieonderzoek bij winkelcentra. In een project bij een winkelcentrum in Roermond kan de politie bepaald verdacht gedrag detecteren, zoals mensen die hun telefoon uitzetten voordat zij het winkelcentrum ingaan. Omdat dit gedrag hooguit verdacht is, maar geen bewijs van misdaad, houdt de politie deze mensen niet aan. In plaats daarvan stuurt zij een bericht naar het winkelcentrum dat er mogelijk dieven rondlopen. De verdachte figuren wordt op deze manier niet persoonlijk onrecht aangedaan. Op structureel niveau verandert het winkelcentrum echter wel meer in een gesurveilleerde ruimte waar mensen elkaar met argwaan bekijken.

Voor AI geldt bovendien dat de middels observatie verkregen gegevens meer vertegenwoordigen dan slechts input voor en output van een digitaal systeem. Data zijn als het ware vormend voor de werking van het systeem, omdat gegevens mede de kwaliteit van risicotaxaties, voorspellingsmodellen en variabelen in het model bepalen. Illustratief is het onderzoek van Van Dijck naar de kwaliteit van het OxRec-algoritme dat de Reclassering gebruikt om rechters te adviseren over het recidiverisico van een verdachte.[729] Met de introductie van dit algoritme lijken de Reclassering Nederland en ook rechters zich volgens Van Dijck te laten leiden door voorspellingen die geregeld niet accuraat zijn, niet beter presteren dan reeds bestaande voorspellingsmodellen en waarbij ongelijke behandeling op basis van ras, sociale klasse of andere sociale ongelijkheden op de loer ligt. Data zijn bovendien niet alleen vormend voor systemen, maar ook voor beleid. Zo wordt inmiddels de transitie van beleidscyclus naar datacyclus bepleit.[730]

Burgers, consumenten en 'anderen' in de gaten houden is voor vrijwel alle bedrijven en overheden en talloze individuele burgers inmiddels staande praktijk geworden. Bedrijven hebben hierop hun verdienmodel gebaseerd, waardoor elke inperking hiervan kosten dan wel derving van potentiële winst met zich meebrengt. En voor overheden geldt dat het verzamelen en verwerken van data, vooral persoonsgegevens, een scala aan mogelijkheden biedt om burgers en bedrijven op hun handelen te controleren.[731] Het verwerken van persoonsgegevens gebeurt bovendien niet alleen op het niveau van individuele personen, maar ook op dat van groepen personen. We hebben eerder al opgemerkt dat AI veelal werkt met data over groepen personen, een praktijk waar bestaande beschermingsmechanismen onvoldoende op zijn toegesneden. Het punt is hier

729 Van Dijck 2020.
730 Van Ginkel en Strijp 2020.
731 WRR 2011: 2016.

niet dat surveillance op voorhand onwenselijk dan wel risicovol is. Waar het om gaat, is een toenemende scheefgroei en disbalans in de verhoudingen tussen burgers, bedrijven en de overheid onderling. Deze scheefgroei betreft allereerst de zeggenschap over en toegankelijkheid, en daarmee bredere beschikbaarheid van de gegevens.[732] We komen daar straks op terug. Problematisch is ook de disbalans in de mate waarin partijen invloed hebben op het verzamelen en de verdere verwerking van de gegevens. Burgers hadden de afgelopen decennia steeds minder zicht en invloed op hetgeen er met hun gegevens gebeurt.[733] Al langer werd gesproken van een *'black box society'*.[734] Met de komst van AI wordt deze black box alleen maar ondoorgrondelijker, ook vanwege het risico dat een vacuüm van adequaat toezicht en rechterlijke controle ontstaat.[735] Want weliswaar kunnen burgers 'elkaar' nu in de gaten houden, maar de onderliggende mechanismen en het achterliggende verdienmodel van de bedrijven die applicaties zoals deurbellen met gezichtsherkenning aanbieden, valt voor hen niet te doorgronden. En met de komst van AI wordt het verzamelen van gegevens in feite steeds ongerichter, in de zin dat het op voorhand niet relevant is om tot selectie over te gaan. De kracht van AI schuilt immers ook in het ontwaren van op voorhand onbekende patronen in grote hoeveelheden data.

Relevant daarbij is dat het niet alleen om een steeds grotere hoeveelheid, ongericht verzamelde, gegevens gaat, maar ook dat het type gegevens een ontwikkeling doormaakt. Waar decennia geleden nog redelijk onschuldige persoonsgegevens werden verzameld en dit veelal in direct contact met burgers gebeurde (door hen te vragen naar bepaalde gegevens), verzamelen talloze slimme apparaten nu zonder dat we er veelal erg in hebben informatie over ons doen en laten. "Op steeds meer plekken in de samenleving worden kenmerken van het lichaam en het gedrag, zoals gezichten, stemmen en emoties, digitaal verzameld en verwerkt. Het gaat om intieme gegevens die gevoelige informatie kunnen blootleggen, bijvoorbeeld over iemands gezondheid, of waarmee iemand op afstand te identificeren is."[736] Aan de hand van verzamelde gegevens over gedrag, emotie en handelen kunnen verzekeraars bijvoorbeeld hun risicoinschatting individualiseren.

Met de komst van AI, en daarmee van de talloze mogelijkheden om bijvoorbeeld gezichtsherkenning in te zetten, krijgt surveillance wederom een nieuwe dimensie. Niet verrassend zien enkele van de toepassingen die in artikel 5 van de EU concept-Verordening AI worden verboden, op surveillance: willekeurige

732 Kop 2020.
733 Moerel en Prins 2016; Solove 2011.
734 Pasquale 2016.
735 De Poorter en Goossens 2019.
736 Gerritsen et al. 2020.

massasurveillance voor rechtshandhaving en *social scoring*, bekend van het sociaal kredietsysteem van de Chinese regering. Een dergelijk voorstel verplaatst terecht de aandacht van AI als zodanig naar het specifieke toepassingsdomein ervan. De vraag is echter of het voldoende is om AI uitsluitend op het niveau van individuele toepassingen en gebruikscontexten te reguleren. Technologiebedrijven gaan tot het uiterste om de aandacht van gebruikers vast te houden, daarmee stijgen immers hun advertentieopbrengsten. Door samen te werken kunnen zij bovendien steeds preciezere gebruikersprofielen maken. Zo krijgt Spotify met behulp van Google Maps inzicht in de muziek die luisteraars in de auto opzetten en kan Google zijn gebruikersprofielen perfectioneren met smaakvoorkeuren. Een dergelijk netwerkeffect zien we overal optreden waar organisaties data verzamelen. De vraag naar de wenselijkheid van een dergelijk sterk gesurveilleerde samenleving is met andere woorden momenteel urgenter dan ooit.

In de discussie over surveillance gaat terecht veel aandacht uit naar de privacy-implicaties en, in het verlengde daarvan, naar kwesties als een verbod op het verzamelen van bepaalde gegevens (bijvoorbeeld biometrische gegevens), transparantie richting burgers en rechten van burgers.[737] Deze discussie zal moeten worden aangevuld en verdiept door aandacht te besteden aan het verdienmodel en de macht van de betrokken bedrijven.[738] De grote nadruk van deze technologiebedrijven op ethisch verantwoorde AI moet om deze reden kritisch worden bezien. Een focus op zelfregulering en de bijbehorende ethiek loopt het risico dat deze de aandacht weg leidt van onderliggende structurele problemen. Door de vraag te stellen of iets op de goede of verkeerde wijze gebeurt, kan de vraag of het überhaupt zou moeten plaatsvinden naar de achtergrond verdwijnen. In haar nieuwe boek maakt Kate Crawford een vergelijkbaar punt. Een nauwe definitie van AI en een abstracte discussie over goed gebruik dienen volgens haar juist de agenda van grote spelers omdat deze vragen over macht, kapitaal en governance achterwege laten.[739] Ethiek is volgens haar dan ook noodzakelijk, maar niet voldoende. De focus moet minder op ethiek liggen en meer op macht.[740]

Scheefgroei

Verzamelen van data, gebruik van data, zeggenschap over data en kwaliteit van data raken met andere woorden steeds meer verweven met fundamentele vraagstukken rondom de inrichting van onze samenleving, hoe burgers de

737 Zie bijvoorbeeld voor de relatie overheid-burger en de inzet van AI daarin: Van Heukelom-Verhage 2020.
738 Häußermann en Lütge 2021; Taylor 2021.
739 Crawford 2021: 9.
740 Crawford 2021: 224.

samenleving zien en het handelen en de positie van individuen in die samenleving.[741] Ook raken deze handelingen aan de positie in de wereld en vooral aan de afhankelijkheid van ons land en Europa van andere delen van de wereld. Zeker nu we vaststellen dat de groeiende hoeveelheid data die beschikbaar komt, in handen is van een zeer beperkt aantal bedrijven van buiten de EU. Daarmee komen we op de tweede ontwikkeling die in verband met de inbedding van AI belangrijk is, namelijk de scheefgroei in aandacht, positie en invloed tussen publieke en private partijen als het gaat om het gebruik van digitale middelen.

Het zijn momenteel vooral private partijen die AI ontwikkelen, gebruiken en in omloop brengen. Dit is te verklaren doordat veel recente doorbraken op het terrein van AI gedaan zijn door het bedrijfsleven. Maar ook doordat overheden wereldwijd tot voor kort bij de regulering van het digitale domein een passieve houding hebben gehad. Hierdoor is sprake van een scheefgroei tussen AI-gebruik in het private domein en AI-gebruik in het publieke domein. Daarmee groeit ook de afhankelijkheid van de overheid van private partijen bij de digitalisering van de publieke sector. Als de overheid AI niet gaat gebruiken, leidt dat enerzijds tot kosten, in de zin van verloren kansen, en anderzijds tot een grotere betrokkenheid van private actoren bij publieke taken, en afhankelijkheid van de overheid en andere publieke organisaties van die partijen. Hierdoor erodeert de democratische verantwoording en kan op termijn ook de situatie ontstaan dat de beleidsruimte van de overheid afneemt.

Dat laatste kan gebeuren wanneer de overheid of organisaties in de zorg of onderwijsinstellingen nog maar van een enkele aanbieder gebruik kunnen maken, die zelf bepaalt welke diensten hij aanbiedt en onder welke voorwaarden. Een voorbeeld hiervan is de huidige discussie over de overstap van de rijksoverheid van Microsoft 365 naar Google Workspace. Hoewel Microsoft herhaaldelijk toezegde de privacy van gebruikers beter te waarborgen, bleef daarover bij de rijksoverheid veel twijfel bestaan. Een overgang naar Google Workspace blijkt vooralsnog problematisch, omdat ook de diensten van dit bedrijf veel privacyrisico's met zich meebrengen. Eenzelfde situatie speelt in het onderwijs, dat gebruik maakt van G Suite for Education, de onderwijsvariant van G Suite Enterprise (onder meer Gmail, Docs en Classroom). Uit een DPIA van de Hogeschool van Amsterdam en de Rijksuniversiteit Groningen komt naar voren dat Google zichzelf ziet als enige verwerkingsverantwoordelijke voor metadata. Google mag dus zelf bepalen voor welk doel zij metadata verzamelen en op welke manier. Ook heeft Google in de privacy-overeenkomsten opgenomen dat zij de voorwaarden rondom metadata eenzijdig mogen aanpassen, zonder de gebruiker daarvoor om toestemming te vragen. Dat betekent

741 Hildebrandt 2018.

dat onderwijsinstellingen die Google G Suite for Education gebruiken, geen of onvoldoende grip houden op wat er met deze gegevens – van zowel personeel als kinderen op basisscholen, middelbare scholen en hoger onderwijs – gebeurt. De Autoriteit Persoonsgegevens adviseert daarom zowel het Rijk als de Nederlandse onderwijsinstellingen Google G Suite niet te gaan gebruiken of het gebruik stop te zetten.

Belangrijk is om vast te stellen dat de risico's die verband houden met de geschetste scheefgroei, niet alleen betrekking hebben op privacy of bijvoorbeeld exclusieve rechten en daarmee op zeggenschap over de AI-applicaties, data en algoritmes. De afhankelijkheid van bedrijven, overigens ook Nederlandse bedrijven, werkt eveneens door in de inhoudelijke invloed die zij in feite hebben bij de vertaling van beleid en regels naar gedigitaliseerde uitvoeringsprocessen. Illustratief zijn de observaties in een rapport van het College voor de Rechten van de Mens over de wijze waarop gemeenten besluiten over de inzet van algoritmes. De auteurs merken op dat veel AI-systemen niet worden ontworpen door gemeenten, maar door bedrijven. "Dit leidt tot een standaardisatie. Dit betekent ook dat alle nationale regels worden geïnterpreteerd door bedrijven en aldus doorwerken in de praktijk van alle gemeenten die deze software gebruiken."[742]

Nederland is zeker niet het enige land dat worstelt met de afhankelijkheid van grote technologiebedrijven. De Duitse overheid heeft een marktanalyse laten uitvoeren, om de afhankelijkheid van individuele softwareaanbieders te verkleinen. De Duitse overheid maakt – net als veel andere nationale overheden – gebruik van Microsoft Cloud, maar heeft besloten dit gebruik af te bouwen, omdat Microsoft vanaf 2026 zijn gebruikers ertoe verplicht gebruik te maken van *software as a service* vanaf de eigen cloud. Dit betekent dat Microsoft vanaf dat moment in feite kan bepalen welke toepassingen gebruikers kunnen gebruiken. Ook eist de Duitse pendant van de Autoriteit Persoonsgegevens op federaal niveau (Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit) van alle Duitse overheidsdiensten dat zij voor het einde van 2021 hun Facebookpagina's afsluiten. Deze pagina's voldoen volgens de autoriteit niet aan de AVG-wetgeving. De beheerders van die pagina's kunnen daarom niet aan hun verantwoordingsplicht voldoen, zoals neergelegd in artikel 5:2 van de AVG.[743] Facebook zelf is naar verluid niet van plan aanpassingen door te voeren. Toch blijken er verschillen tussen landen onderling in de mate van afhankelijkheid. Zo liet een EU-landenvergelijking van universitaire ICT-diensten zien dat ons land zich in veel sterkere mate afhankelijk heeft

742 Choi et al. 2021: 5.
743 Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit, 20 mei 2019.

gemaakt van Amerikaanse cloudaanbieders dan de meerderheid van de andere EU-lidstaten, waaronder Duitsland en Frankrijk.[744]

Eenzelfde scheefgroei ligt bij AI in het verschiet, zeker wanneer deze als dienst aangeboden gaat worden. Afhankelijkheden zijn er zowel op het terrein van AI zelf als op dat van de ondersteunende technologie die nodig is om AI te kunnen gebruiken (zie tekstbox 7.7). Ook hier zullen gebruikers waarschijnlijk in het begin meerdere keuzes aangeboden krijgen. Gaandeweg echter kunnen de grote technologiebedrijven hier eveneens beslissen dat ze AI-toepassingen alleen nog aanbieden wanneer gebruikers tevens hun data bij hen onderbrengen, of dat ze AI enkel aanbieden als onderdeel van een groter totaalpakket.

Tekstbox 7.7 – AI en afhankelijkheid van buitenlandse aanbieders

Grote afhankelijkheden van buitenlandse aanbieders zijn er zowel op het terrein van AI zelf als op dat van de ondersteunende technologie.[745] Nederland is goed in onderzoek en ontwikkeling van AI maar kent afhankelijkheden waar het gaat om de toegang tot aan AI gerelateerde diensten. Hierbij gaat het met name om de toegang tot software-pakketten en onlinediensten van bibliotheken, die voor het overgrote deel door de grote technologiebedrijven worden geleverd. Dit geldt behalve voor betaalde software ook voor gratis en open source software. Voorbeelden zijn modellen voor beeldherkenning die tegen betaling toegankelijk gemaakt door bijvoorbeeld Google. Of AI-management-tools en -diensten zoals de Machine Learning Engine op Google Cloud, Azure Machine Learning Studio op Microsoft Azure, Einstein op de Salesforce cloud of IBM Watson ML. Hoe meer dergelijke software wordt ingebed in Nederlandse AI-producten en -diensten, hoe minder controle er is over de waarborging van publieke waarden. Ook is de vraag in hoeverre deze pakketten open source blijven.

Op het vlak van de ondersteunende technologie valt de grote afhankelijkheid van diensten en producten van buitenlandse cloudaanbieders op, die een risico vormt voor de gehele waardeketen van AI-toepassingen. Nederland is in Europa de vierde grootste markt voor cloudinfrastructuur in Europa. Deze markt wordt echter gedomineerd door buitenlandse partijen. Zo staan Amazon, Microsoft en Google in de top drie en komt KPN daarna, op de vierde plek.

744 Fiebig et al. 2021.
745 Gebaseerd op TNO 2021.

Er zijn twee reden voor dit scenario. De eerste is – zoals hierboven ook bleek – dat dergelijke lock-ineffecten nu al aan de orde van de dag zijn. Dit heeft ermee te maken dat bedrijven van netwerkeffecten profiteren. Hoe meer gebruikers ze hebben, des te meer data ze verwerven, en des te beter ze hun producten kunnen optimaliseren. Met andere woorden: het is in lijn met het bedrijfsbelang om gebruikers vast te houden. Het aanbieden en later verplicht stellen van integrale dienstenpakketten is een manier om dit te bereiken, een andere optie is om toepassingen incompatibel te maken met die van andere aanbieders.

De tweede reden is dat juist omdat deze partijen de beschikking hebben over grote hoeveelheden data, ze goed gepositioneerd zijn om AI te ontwikkelen en daarin ook op grote schaal investeren. Het afgelopen decennium hebben honderden overnames van AI-bedrijven plaatsgevonden; de grootste opkopers waren achtereenvolgend Apple (20), Google (14), Microsoft (10), Facebook (8) en Amazon (7).[746] Een derde factor die we daarom bespreken is machtsconcentratie.

Machtsconcentratie

Bij de factor machtsconcentratie gaat het om de invloedrijke positie van de grote Amerikaanse technologiebedrijven, die tevens de AI-ontwikkeling domineren.[747] Daaronder bevinden zich dus ook de bovengenoemde bedrijven waarmee de overheid werkt. Naarmate AI verder doordringt in de samenleving, groeit hun invloed op tal van maatschappelijke terreinen, waaronder ook politieke processen en de democratie.[748] Hiervoor waarschuwt bijvoorbeeld Paul Nemitz, adviseur van de Europese Commissie en lid van de commissie voor data-ethiek van de Duitse regering. Nemitz hanteert het standpunt dat de grote impact die AI op de samenleving heeft, vereist dat daarover democratisch gelegitimeerde besluiten worden genomen. Het is hier dus niet langer de technologie zelf die resulteert in de roep om nieuwe spelregels, maar de toename in het gebruik van AI die een beperkt aantal bedrijven in de positie brengt om een te grote stempel op de samenleving te drukken. Dat is met name problematisch als de op AI gebaseerde diensten van deze bedrijven tot de maatschappelijke infrastructuur gaan behoren en de vraag voorligt of deze diensten als publiek goed moeten worden beschouwd.[749]

We kennen de problematiek van machtsconcentratie ook van de olie-industrie en de elektriciteitsbedrijven. Aanvankelijk zeer innovatief, ontwikkelden zij zich gaandeweg tot monopolisten, waarna ze door de overheid werden

746 Zie: Nemitz en Pfeffer 2020: 84.
747 Poon 2016; Moore en Tambini 2018.
748 Fukuyama 2021.
749 CPB 2019.

opengebroken, in stukken opgedeeld of tot algemeen nutsbedrijf omgevormd. Zo waren vlak voor het uitbreken van de Eerste Wereldoorlog GE en Westinghouse uit de VS en Siemens en AEG uit Duitsland de grootste bedrijven ter wereld, gedreven door fusies om de schaal en de toegang tot kapitaal te vergroten. Onderling besloten zij zelfs om de wereldmarkt te verdelen wat betreft het produceren en exporteren van elektriciteitstechnologie en elektrische machines.[750]

Onderzoekscommissies van overheden in de VS, Europa en het Verenigd Koninkrijk hebben de afgelopen jaren hun pijlen gericht op de grote technologiebedrijven, die ze betichten van machtsmisbruik. Ze signaleren dat er niet langer competitie is op de markt, maar slechts competitie om de markt, met als resultaat minder innovatie, minder keuze voor consumenten, ondermijning van privacyrechten, verzwakking van de pers en een verzwakking van de democratie. Hun analyse van de positie van deze bedrijven is daarbij opvallend eensluidend.[751]

De gerechtelijke commissie van het Amerikaanse Huis van Afgevaardigden bracht begin oktober 2020 een rapport uit over Apple, Facebook, Google en Amazon. De conclusie van het rapport was dat de bedrijven poortwachters zijn van een bepaald distributiekanaal en deze macht misbruiken om anderen de pas af te snijden en zo de machtigste te blijven. Volgens de commissie is rondom deze bedrijven sprake van een *'kill zone'* – concurrenten overleven alleen als ze ver uit de buurt blijven. Bovendien maken de bedrijven misbruik van hun rol als bemiddelaar om hun eigen dominantie te versterken. Zo gebruikt Amazon data over bedrijven die van de cloud gebruikmaken, om eigen concurrerende producten aan te bieden.

Het rapport besteedt ook aandacht aan AI. Het constateert bijvoorbeeld dat er bij stemassistenten een duidelijk netwerkeffect is. Alle toepassingen met algoritmes leren en worden beter naarmate ze meer gebruikt worden. Gebruikersaantallen zijn met andere woorden van doorslaggevend belang in termen van AI. De toegang tot grote hoeveelheden data in combinatie met AI biedt deze bedrijven bovendien de kans om andere markten te betreden waar data een rol spelen. Zij profiteren daarvan bijvoorbeeld nu al op de markt voor 'slimme' apparaten. Apple en Amazon (Alexa) verkopen voor dat doel bijvoorbeeld slimme speakers, die bij voorkeur alleen de eigen diensten openen

750 Bakker en Korsten 2021: 34.
751 Zie hiervoor: Lancieri en Sakowski 2020. In deze studie analyseren de auteurs 22 onderzoeken van experts en mededingingsautoriteiten naar concurrentie in digitale markten.

of aanprijzen, voor dumpprijzen.[752] Als kopen via de speaker in de toekomst de standaard wordt, hebben zij deze markt op dat moment al stevig in handen.

Om meer grip op de geschetste ontwikkelingen te krijgen circuleren inmiddels verscheidene voorstellen. Tegelijkertijd toont de discussie in feite ook de onmacht van het huidige instrumentarium. Privacywetgeving, bijvoorbeeld, is het standaardantwoord als het gaat om de bescherming van persoonsgegevens. Maar er zijn intussen grote twijfels met betrekking tot de vraag of dit instrumentarium voldoende en voldoende effectief is.[753] Ook is er kritiek op de focus van het mededingingsrecht, die te veel is komen te liggen op lage prijzen als voornaamste indicator voor consumentenwelvaart.[754] Deze focus is slecht toepasbaar op de technologiebedrijven, omdat ze (een deel van) hun diensten overwegend gratis aanbieden, en vaak meerdere markten tegelijk bedienen. Een heroverweging van het mededingingsrecht en de onderliggende doelen daarvan is een optie die daarom veel wordt genoemd.[755] Het Amerikaanse rapport suggereert onder andere om de bedrijven op te breken. Wat pleit voor de (dreiging van) opsplitsing van de grote technologiebedrijven, is dat deze zorgt voor een enorme positieve dynamiek op de markt. Voorbeelden zijn IBM (splitsing hardware en software), AT&T (destijds het grootste bedrijf ter wereld, opgesplitst in acht kleinere bedrijven) en Microsoft.[756]

Ook de Europese Commissie is op dit terrein actief.[757] De Commissie voert bovendien al jaren rechtszaken tegen technologiebedrijven die zich niet houden aan de Europese wet- en regelgeving en deelt daarbij steeds hogere boetes uit. Ook zijn er sinds eind 2020 twee nieuwe wetsvoorstellen, de Digital Service Act en de Digital Markets Act. De eerste, 'Proposal for a Regulation on a Single Market For Digital Services (DSA)', richt zich op grote platforms om illegale en schadelijke inhoud snel te verwijderen en gebruikers inzicht te geven in de manier waarop algoritmes met aanbevelingen werken. De Digital Service Act heeft tot doel aansprakelijkheids- en veiligheidsregels voor digitale platforms, diensten en producten te verbeteren. De grootste platforms komen daarbij onder ingrijpend systeemtoezicht te staan.

Met de Digital Markets Act komt de Commissie bovendien met extra regels bovenop het mededingingsrecht. De grote technologieplatforms worden in dit

752 Committee on the Judicary 2020: 124-125, 307-312, 377.
753 Purtova 2018: 40-81.
754 Gerbrandy en Custers 2018
755 Zie over de verschillende voorstellen: Kohlen et al. 2021.
756 Wu 2020: hoofdstuk 5.
757 Crémer et al. 2019; Kohlen et al. 2021.

wetsvoorstel getypeerd als 'digitale poortwachters'.[758] Hiermee wordt enerzijds de omvang van deze bedrijven erkend en anderzijds aangeduid welke cruciale rol zij vervullen in de samenleving. Deze langverwachte karakterisering zou wel eens een einde kunnen maken aan de situatie waarin de grote technologiebedrijven in staat zijn zich te onttrekken aan wetgeving. Het argument dat ze een bijzondere categorie bedrijven vormen en allerhande wetgeving daarom niet van toepassing is, gaat met het voorstel van de Commissie tot het verleden behoren. Welke vorm dit voorstel uiteindelijk ook krijgt, het 'digitale mededingingsrecht' gaat er fundamenteel door veranderen.[759]

Ook in Nederland loopt de discussie over het mededingingsrecht al enige jaren. Zo presenteerde het voormalige kamerlid voor D66 Verhoeven in 2019 een initiatiefnota om te komen tot een modernisering van de mededingingsregels, een plek voor data in de Europese en nationale mededingingsregels en nieuwe criteria voor het afbakenen van de omvang van de digitale markt en om het marktaandeel van een bedrijf vast te stellen.[760] Ook in brieven aan de Tweede Kamer besteden diverse ministers aandacht aan de geschetste ontwikkelingen.[761] Belangrijk in dit verband is ook de discussienotitie *Toekomstbestendigheid mededingingsbeleid in relatie tot online platforms* van de staatssecretaris van Economische Zaken en Klimaat uit 2019. Deze behandelt ook een aantal specifieke ontwikkelingen rondom de inzet van algoritmes en kartelvorming: "Doordat de voorkeuren en inkomens van consumenten met behulp van data en algoritmes steeds preciezer in te schatten zijn, kan geïndividualiseerde prijsdiscriminatie zich gaan voordoen."[762] Zelflerende algoritmes zouden in de toekomst zelfs zonder menselijke tussenkomst kartels kunnen gaan vormen, aldus de notitie. In aansluiting hierop is een conceptwetsvoorstel uitgebracht voor de modernisering en betere handhaving van consumentenbescherming.

Inmiddels heeft ons land samen met Duitsland en Frankrijk bovendien een voorstel gepubliceerd dat moet voorzien in een beoordeling door een EU-toezichthouder van alle fusies en overnames door grote digitale platforms met een poortwachtersfunctie.[763] De drie landen doen dit voorstel ter aanvulling, en daarmee versterking, van het reeds in de EU Digital Markets Act voorgestelde toezicht en andere maatregelen. Denk hierbij aan de verplichting tot het delen

758 Zie onder meer het voorstel voor de EU Digital Markets Act (DMA).
759 Chavannes 2021: 17-20.
760 Kamerstukken 2018/19, 35 134, nr. 2.
761 Kamerstukken II 2019/20, 26643, nr. 672.
762 Bijlage bij de brief van staatssecretaris van Economische Zaken en Klimaat, 17 mei 2019: 6.
763 Voorstel van 26 mei 2021 (Le Maire, Altmaier en Keijzer, 26 mei 2021). Vergelijk *Considerations of France and the Netherlands regarding intervention on platforms with a gatekeeper position*, gepubliceerd op 15 oktober 2020.

van data, interoperabiliteit en het verbod op het bevoordelen in rangordes van eigen producten of diensten door deze digitale poortwachters.

Hiernaast circuleren er ook andere voorstellen voor het kleiner of minder invloedrijk maken van de grote technologiebedrijven, bijvoorbeeld door ze te dwingen hun diensten en data beschikbaar te stellen voor anderen.[764] De discussie gaat daarbij over interoperabiliteit en platformneutraliteit, waarbij de laatste is gebaseerd op het reeds bestaande principe van netneutraliteit dat internetproviders verbiedt om content verschillend te beprijzen. Een relatief nieuw element in de discussie is de benadering waarbij gepleit wordt voor een combinatie van regulering en technologische oplossingen. Deze benadering richt zich op zogenoemde *middleware*, die voor concurrentie moet zorgen bovenop de bestaande platforms.[765]

De bedrijven die deze middleware verzorgen, richten zich op het editen van nieuws en informatie, hebben daar eigen algoritmes voor en kunnen daar ook een eigen profiel in ontwikkelen. Enerzijds maakt dit het voor gebruikers mogelijk om te kiezen voor verschillende informatiekanalen en anderzijds voorkomt het daarmee ook dat foutieve informatie of nepnieuws snel een enorme schaal krijgt, doordat de algoritmes van een enkel platform daarmee aan de haal gaan. Dit voorstel richt zich dan ook met name op de problemen die platformbedrijven momenteel veroorzaken, gekoppeld aan de verspreiding van nepnieuws en het filteren van illegale content. Het is echter lastig om een dergelijke structurele aanpassing in de digitale infrastructuur te implementeren. Er is een nieuw verdienmodel voor nodig, platforms moeten de toepassing willen ondersteunen en er moet een technisch raamwerk komen dat tegelijkertijd aansluit op de architectuur van de verschillende platforms én een gevarieerde markt mogelijk maakt.

Hoe precies de voorstellen van de Europese Commissie en andere partijen hun beslag zullen krijgen, is vooralsnog onduidelijk. Maar helder is dat ze de inbedding van AI in de samenleving sterk zullen beïnvloeden. De overheid heeft dus niet alleen werk te verrichten als het gaat om de vragen die AI oproept ten aanzien van de huidige kaders, maar zal tegelijkertijd moeten investeren in manieren om de omgang met AI zodanig te structureren dat publieke waarden daarbij eveneens voor de verdere toekomst voldoende gewaarborgd blijven. De inbedding van AI in de samenleving is daarmee vooral een 'inrichtingsvraagstuk' dat betrekking heeft op de bredere digitale leefomgeving. Op het terrein van dat inrichtingsvraagstuk zal de overheid echt onverwijld stappen moeten nemen.

764 Graef en Prüfer 2021.
765 Fukuyama et al. 2021.

Dit om te voorkomen dat de kaarten straks grotendeels zijn geschud omdat andere partijen het voortouw hebben genomen bij hoe en voor welke doelen AI wordt ingezet. Of doordat burgers AI zijn gaan wantrouwen of zelfs afwijzen.

Kernpunten – Regulering van de inbedding van AI in de digitale leefomgeving

- Onherroepelijk zal de rol van de overheid bij de regulering van AI toenemen naarmate de technologie breder verspreid raakt in de samenleving en de noodzaak ontstaat om verdere eisen te stellen.
- Timing is van groot belang bij overheidsinterventie. Wanneer de overheid te lang wacht met interveniëren, kan AI in de samenleving ingebed raken op een wijze die strijdig is met publieke waarden of die onvoldoende dienstbaar is aan die waarden.
- Een systeemtechnologie als AI vraagt om een wetgevingsagenda met daarop, behalve kwesties rond de technologie zelf en het gebruik ervan, vooral ook de bredere maatschappelijke effecten van AI.
- Massasurveillance, de grote afhankelijkheid van private partijen en machtsconcentratie zetten publieke waarden bij de maatschappelijke inbedding van AI onder druk en behoeven acute actie van de overheid

7.3 Tot slot

De opgave die in dit hoofdstuk centraal staat, is die van regulering door de overheid. Bij deze opgave hebben we twee dimensies onderscheiden die samenhangen met het systeemkarakter van AI. De eerste dimensie betreft de alomtegenwoordigheid van AI. Deze dimensie impliceert dat AI op een groot aantal terreinen nieuwe of aangepaste kaders zal vereisen.

Op dit moment staat de inbedding van AI in de samenleving nog overwegend in de kinderschoenen. Daaraan zou de conclusie verbonden kunnen worden dat de overheid geen actie behoeft te ondernemen en een afwachtende houding kan aannemen. Gezien de grote impact die AI naar verwachting zal hebben, is dat echter onwenselijk. Wil de overheid in staat blijven om ook op langere termijn effectief en betekenisvol te interveniëren, zeker met het oog op publieke waarden, dan moet ze waakzaam zijn en zich nu al voorbereiden op de sterkere rol die onherroepelijk op haar afkomt. Dit proces is inmiddels op een aantal terreinen op gang gekomen, zowel in Nederland als in de Europese Unie. Belangrijk is dat de overheid zich daarbij bewust is van de verschillende kwesties die spelen bij de regulering van AI. Ook zal zij structureel moeten investeren in het vergaren en bijeenbrengen van signalen over de kansen en risico's waarmee de inbedding van AI in de samenleving gepaard gaat. Zijn dergelijke signalen niet

voorhanden, dan kunnen te zijner tijd niet dan wel onvoldoende tijdig de juiste aanpassingen of nieuwe regels worden opgesteld.

De tweede conclusie van dit hoofdstuk is dat naarmate AI meer ingebed raakt in de samenleving, het werk voor de overheid alleen maar zal toenemen. Bij die grotere rol verandert bovendien de aard van de vraagstukken die er voor de overheid liggen, omdat de toename in het gebruik van AI tot de komst van tweede- en derde-ordeproblemen leidt. Het is dan niet meer uitsluitend de technologie als zodanig die vragen oproept, maar de omvang van het gebruik ervan en de schaal van de effecten die het probleem zijn. De overheid zal daarom ook actief moeten sturen op de bredere digitale leefomgeving, waarin AI uiteindelijk ingebed raakt. Zoals zij dat eerder deed op andere terreinen waar ontwikkelingen de samenleving in fundamentele zin raken en (her)ordenen. Dit vereist dat de overheid de komende periode inzet op het samenbrengen van losse beleidsdossiers en ontwikkelingen en deze beziet als onderdelen van een omvattender inrichtingsvraagstuk.

Belangrijk daarbij is om vast te stellen dat AI terecht komt in een samenleving waarbinnen reeds op grote schaal data worden verzameld, met name private partijen digitale producten, diensten en infrastructuren aanbieden en de voornaamste ontwikkelaars van AI een dominante positie hebben in de wereldwijde interneteconomie. Deze context is van grote invloed op de wijze waarop AI uiteindelijk in de samenleving gebruikt zal gaan worden, door wie en voor welke doelen. Wil de overheid greep blijven houden op dit proces, dan moet zij nu ingrijpen. Wachten is hier niet alleen onwenselijk, maar bovendien niet noodzakelijk. Er liggen inmiddels al talloze onderzoeksrapporten, andere documenten en plannen klaar om dit ingrijpen te onderbouwen en richting te geven. Waar het nu vooral op aankomt, is deze plannen voortvarend om te zetten in regulerend handelen.

8. Positionering

De laatste opgave die wij onderscheiden, is het *internationaal positioneren* van ons land ten aanzien van AI. Deze opgave heeft een iets ander karakter dan de andere vier. Deze doelt namelijk op een niveau of een speelveld waar alle voorgaande opgaven ook ten dele betrekking op hebben. Mythen over AI zijn er ook binnen het internationale speelveld en het adresseren daarvan vindt plaats vanuit internationale media, bedrijven en onderzoeksinstituten. Verschillende vraagstukken rondom contextualisering hebben een internationale dimensie. Het technische ecosysteem omvat bijvoorbeeld de mondiale discussie rondom de uitrol van 5G en Europese ambities als Gaia-X om een gezamenlijke data-infrastructuur te ontwikkelen. Het engageren van stakeholders kan lokaal of nationaal, maar heeft ook een internationale dimensie. De rol van wereldwijde wetenschappelijke associaties en ngo's is daar een voorbeeld van. Regulering vindt voor een belangrijk deel op internationaal niveau plaats. Denk hierbij aan internationale verdragen over gevaarlijke toepassingen van technologie, ethische richtlijnen of de ambities van de EU om AI te reguleren.

Ondanks dat er een sterke verwevenheid is met de eerder besproken opgaven, is het zinvol om de internationale dimensie van de inbedding van AI apart te behandelen. In de eerste plaats omdat het hierbij om een specifiek type actoren gaat. In internationale gremia wordt Nederland door bepaalde partijen gerepresenteerd die onderhandelen en samenwerken met actoren uit het buitenland. Dit zijn niet alleen statelijke actoren, maar ook internationale organisaties, multinationals en zelfs individuen.

Er is nog een tweede reden om apart naar het internationale veld te kijken, namelijk vanwege twee vraagstukken die specifiek op dit niveau spelen. Het eerste heeft te maken met de concurrentiekracht of het verdienvermogen van Nederland. Vragen over waar het competitief voordeel van Nederland ligt, of onze positie versterkt kan worden en hoe dit zich verhoudt tot de capaciteiten van andere landen, vallen hieronder. Het tweede voor ons hier relevante vraagstuk, met bij uitstek een internationale dimensie, gaat over veiligheid.[766] De meest extreme activiteit op dit gebied is oorlogsvoering. Nieuwe technologieën hebben grote invloed op de manier waarop oorlog gevoerd wordt en hoe daarbij successen kunnen worden geboekt. Bij AI gaat het in deze context vaak over autonome wapens, al is de invloed van AI op oorlogsvoering veel breder dan dat. Veiligheid speelt daarnaast in minder extreme situaties en gaat dan over zaken

766 Wij zeggen hier 'bij uitstek', omdat veiligheid natuurlijk ook een binnenlandse aangelegenheid is en omdat die twee tonelen steeds meer met elkaar verweven raken. Zie: WRR 2017.

als buitenlandse beïnvloeding en de export van ideologieën, maar ook over sabotage en industriële spionage. Dergelijke vraagstukken spelen niet alleen tussen landen met vijandige relaties, maar zelfs tussen bondgenoten. Denk hierbij aan de dimensie van *flow security* die gaat over de veiligheid van allerlei goederenstromen van voedsel en medicijnen, maar ook van data, kapitaal en mensen.[767]

De vraagstukken ten aanzien van verdienvermogen en veiligheid kunnen ook met elkaar verweven zijn. In de discussie rondom 5G-technologie wordt het Chinese bedrijf Huawei niet alleen gezien als een economische concurrent, maar vooral als een veiligheidsrisico. In het Amerikaans-Chinese handelsconflict gaan economische argumenten samen met vragen over nationale veiligheid.[768] Ook de discussie in Europa over de macht van de Amerikaanse Big Tech-bedrijven ging aanvankelijk over competitiviteit, maar wordt meer en meer geduid vanuit de vraag naar strategische autonomie en digitale soevereiniteit.[769] Een groeiende literatuur over 'geo-economie' benadrukt de sterke verbinding tussen economie en verdienvermogen enerzijds en geopolitiek en veiligheid anderzijds.[770] In dit hoofdstuk onderzoeken we de twee vraagstukken apart, waarbij we gaandeweg steeds op verbindingen tussen de twee wijzen. Figuur 8.1 laat zien hoe de thema's uit dit hoofdstuk zich verhouden tot de vraagstukken van verdienvermogen, veiligheid en de tussenliggende geo-economie.

Een laatste reden om apart bij het internationale veld stil te staan, is de vraag naar het niveau waarop sommige uitdagingen geadresseerd moeten worden. Een aantal gebieden, zoals het mondiale financiële systeem, is internationaal zo verweven dat het nationale toneel niet volstaat om bepaalde uitdagingen het hoofd te bieden. Op veel gebieden is de Europese Unie het niveau van regels en afspraken geworden, maar op andere gebieden spelen mondiale organisaties als de Verenigde Naties (VN) of allianties als de NAVO een belangrijkere rol. Aparte aandacht voor het internationale veld is dus ook belangrijk om antwoord te kunnen geven op de vraag op welk niveau bepaalde vraagstukken het beste geadresseerd kunnen worden.

---

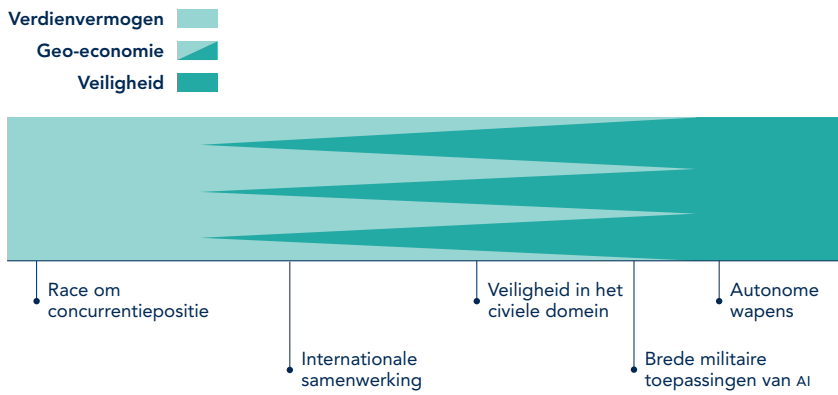
767 WRR 2017.

768 Daniel Drezner laat zien hoe onder Donald Trump de economie meer is ingezet als wapen in strategische vraagstukken (Drezner 2019).

769 Het concept van strategische autonomie werd tot recent vooral in Frankrijk gebruikt voor het militaire domein en in India ten tijde van de Koude Oorlog. De afgelopen jaren passen allerlei Europese politici van Emanuel Macron tot Peter Altmaier het toe op het domein van digitalisering (Timmers 2019). Zie in ons land ook het advies hiervan van de Cyber Security Raad (2021).

770 De term geo-economie is gemunt in Luttwak 1990. Een recent overzicht van de literatuur sindsdien is te vinden in Scholvin en Wigell 2018.

Figuur 8.1 Thema's op het gebied van verdienvermogen, veiligheid en de tussenliggende geo-economie



De centrale vraag bij deze opgave is: *Hoe verhouden wij ons als Nederland internationaal?* We gaan eerst in op de internationale positionering in relatie tot het verdienvermogen van Nederland (paragraaf 8.1). Daarbij kijken wij specifiek naar nationale AI-capaciteiten, het fenomeen van AI-strategieën en het vaak daarmee gepaard gaande idee dat zich een mondiale AI-race afspeelt. Vervolgens bespreken wij internationale positionering in relatie tot veiligheid (paragraaf 8.2). We kijken naar autonome wapens, maar ook naar andere manieren waarop AI invloed kan hebben op oorlogsvoering. Ook behandelen we bredere veiligheidsvraagstukken tussen landen en gaan we in op de opkomst van een 'digitale dictatuur'.

8.1 AI en de concurrentiepositie van Nederland

AI-capaciteiten

Net als eerdere technologische revoluties is AI van invloed op de concurrentiepositie van landen en zijn er grote verwachtingen over de economische waarde die de technologie kan creëren. In hoofdstuk 2 zagen we al dat PwC een bijdrage van AI voorspelt van 15,7 biljoen dollar aan de wereldeconomie in 2030.[771] Hoe ziet het internationale economische speelveld er nu uit? Zoals uit eerdere hoofdstukken is gebleken, is AI een complex fenomeen met verschillende dimensies. Het is daarom niet mogelijk om het speelveld met een enkele maatstaf in kaart te brengen. Er zijn indices voor de economische waarde van AI-activiteiten, het aantal AI-spelers per land, het aantal patenten en er is

771 Rao en Verweij 2017.

inmiddels ook een AI-index.[772] Allemaal laten ze echter slechts een deel van het plaatje zien. De auteurs Jeffrey Ding en Kai-Fu Lee hebben gepoogd tot een indeling te komen waarmee de AI-capaciteiten van een land zijn in te schatten.[773] Als we hun benaderingen combineren komen we tot vijf variabelen:

1. **kwaliteit van fundamenteel onderzoek**
2. **beschikbare data**
3. **vereiste hardware**
4. **dynamisch bedrijfsleven dat AI kan commercialiseren**
5. **stimulerende overheid**

De eerste drie variabelen zijn we in hoofdstuk 5 al tegengekomen bij het technische ecosysteem, toen het ging over de benodigdheden om AI te laten functioneren. De as van het bedrijfsleven omvat zowel grote technologiebedrijven als een innovatieve startup-cultuur en de AI-investeringen van grote bedrijven in andere sectoren. Een overheid kan stimulerend zijn door te investeren, maar ook door wetgeving te ontwikkelen die duidelijkheid creëert of zelfs speciale experimenteerruimte.

Als we deze vijf assen als uitgangspunt nemen, wordt duidelijk dat de VS en China de twee grote wereldleiders zijn. Beide landen zijn op alle dimensies sterk. Beide landen beschikken door hun omvang en relatief lichte wetgeving over grote hoeveelheden data. Ook hebben ze grote diverse technologiebedrijven: in het ene geval Big Tech uit Silicon Valley met bedrijven als Alphabet, Amazon, Facebook, Microsoft en Apple, en in het andere geval giganten als Baidu, Alibaba, Tencent (afgekort 'BAT') en Huawei. Daarnaast groeien in beide landen gespecialiseerde bedrijven die AI gebruiken, ook snel uit tot grote ondernemingen, zoals Uber en Netflix enerzijds en Bytedance en Hikvision anderzijds.

De adoptie van AI op consumentenplatformen in China is erg hoog; spraakherkenningssoftware wordt veel gebruikt en consumenten kunnen betalingen doen met hun gezicht.[774] Op het gebied van fundamenteel onderzoek leidt de VS, maar China is hard op weg om dat verschil te verkleinen.[775] Een studie naar citaties van onderzoeksinstituten tussen 2012 en 2016 liet zien dat China

772 Stanford University publiceert jaarlijks het AI-Index Report over de mondiale AI-trends.
773 Ding 2018; Lee 2018.
774 Lee 2018: 118.
775 In de laatste AI-index van Stanford staat dat China sinds 2017 de EU voorbij is in het aantal wetenschappelijke publicaties én het percentage van het totaal aantal publicaties (en dat het de VS al in 2008 voorbij was). Als het gaat om citaties van wetenschappelijke artikelen, is China in 2016 de EU voorbijgegaan en in 2019 ook de VS. Maar als het gaat om het 'gewicht' van de verwijzingen, staan zowel de VS als de EU nog steeds boven China (Zhang et al. 2021: 18-30).

nummer twee is achter de VS en dat in de categorie elite-instituten Tsinghua University zelfs hoger scoorde dan Stanford University in het totaal aantal AI-citaties.[776] Een andere studie naar academische papers die gepresenteerd werden op grote AI-conferenties, liet een daling zien van het percentage auteurs van Amerikaanse instituten van 41 procent in 2012 naar 34 procent in 2017 en een stijging van het percentage Chinese auteurs van 10 naar 24 procent.[777]

Ook anekdotische informatie bevestigt deze trend. De Chinese startup Face++ domineerde in 2017 een internationale wedstrijd in beeldherkenning en versloeg daarbij teams van Google, Microsoft en Facebook. Op het gebied van spraakherkenning is het Chinese iFlyTek het Amerikaanse Nuance gepasseerd en beide bedrijven behoren nu tot de wereldtop. Voormalig CEO van Google Eric Schmidt waarschuwde in 2017 voor zelfgenoegzaamheid tegenover Chinese AI-vaardigheden en voorspelde dat het land de VS vijf jaar later zou evenaren.[778]

Jeffrey Ding en Kai-Fu Lee schatten de relatieve verhouding tussen de twee landen echter verschillend in. Voor Ding is de VS met enige afstand de wereldleider en zal dat voorlopig blijven. Lee zet zijn kaarten in op China. Het verschil heeft te maken met het gewicht dat beiden toekennen aan de verschillende assen. Ding laat zien dat de VS vooral op het gebied van hardware (specialistische chips) een zeer sterke positie heeft. China is hiervan afhankelijk en tegelijkertijd wordt de VS restrictiever in het delen van die technologie. Lee wijst daarentegen op het feit dat China veel meer data heeft en daar relatief ongehinderd mee kan werken, wat volgens hem in de toepassingsfase van AI allesbeslissend zal zijn. Daarnaast wijst hij op de rol van de overheid. Hoewel de VS ook een AI-strategie heeft, is er geen enkele strategie zo ambitieus als die van China.

In juli 2017 werd het *'New Generation Artificial Intelligence Development Plan'* (AIDP) gepubliceerd. Daarin staan de precieze doelen van China voor de aard en de omvang van AI in de komende jaren. In het AIDP formuleert China de ambitie om in 2020 op vergelijkbaar niveau te staan met de meest geavanceerde landen op het gebied van AI, in 2025 wereldleider te zijn in sommige gebieden van AI en in 2030 het primaire AI-innovatiecentrum van de wereld te zijn. Naast deze ambities geeft het plan lokale overheden een signaal om hun eigen plannen en fondsen op te zetten. Ook benoemt het plan de belangrijkste beleidsinstrumenten om de gestelde doelen te bereiken.

---

776 Een andere studie onderzocht citaties in de top 100 van AI-tijdschriften en -conferenties tussen 2006 en 2015. Daaruit bleek dat het aandeel papers van auteurs met Chinese namen toenam van 23,2 naar 42,8 procent (Lee 2018: 89).

777 Leung 2019: 250.

778 Lee 2018: 90, 105.

Opvallend is de nadruk op het bepalen van technische standaarden: maar liefst 24 keer wordt dat in het AIDP genoemd. Op de rol van standaardisatie komen we later in dit hoofdstuk terug. Het AIDP benadrukt verder het belang van internationale samenwerking op het gebied van regulering en ethische normen voor AI.[779] AI wordt in China inmiddels op allerlei domeinen gebruikt. Het technologieplatform Tencent heeft bijvoorbeeld een zorgprogramma gelanceerd, Miying, om medische professionals te assisteren bij diagnoses. De politie gebruikt gezichtsherkenningssoftware, maar ook software om lichaamshoudingen te analyseren. Fondsen worden gebruikt voor toepassingen in het onderwijs en het zakenleven. In de stad Hangzhou bouwt Alibaba 'City Brain', een AI-systeem voor verkeersmanagement en een betere respons van hulpdiensten.[780]

Het is geen uitgemaakte zaak of de inschatting van Ding of Lee accurater is. Wel is duidelijk dat deze twee landen in mondiaal perspectief de twee 'AI-supermachten' zijn. Hoe zit het met de rest van de wereld? Als geheel staat de EU er niet slecht voor. Op het gebied van fundamenteel onderzoek loopt de EU voor op China en is ze vergelijkbaar met de VS. De beschikbaarheid van data is minder goed vanwege nationale diversiteit binnen de EU, maar ook door strengere wetgeving.[781] Op het gebied van hardware staat de EU er goed voor. Europese landen zijn voor specialistische chips afhankelijk van de VS, maar hebben daar door vriendschappelijke relaties vooralsnog vrij toegang toe. Verder boekt de EU vooruitgang op het gebied van een stimulerende overheid. Naast nationale AI-strategieën is er ook een strategie op Europees niveau.[782] De totale investeringen zijn vooralsnog relatief klein, maar in de EU groeit het momentum om AI als strategische technologie te stimuleren. De COVID-19-pandemie lijkt bovendien aan dit momentum bij te dragen, Zo gaat 20 procent van de 670 miljard euro aan budget dat is uitgetrokken voor het EU-herstelplan, naar de bredere ontwikkeling van digitalisering, maar AI zal hier onherroepelijk van profiteren. Dat geldt ook voor de middelen die zijn gemoeid met het EU4Health programma, de Connecting Europe Facility – Digital dat de infrastructuur moet gaan bekostigen, en het Digital Europe-programma.[783]

De grootste zwakte van de EU ligt in het bedrijfsklimaat rondom AI. Bedrijven in andere sectoren zoals infrastructuur en energie ontwikkelen wel degelijk hun AI-capaciteiten. Ook zijn er technologiebedrijven als SAP, Dassault, ASML en

779 Ding 2019: 43-44.
780 Creemers 2019: 130.
781 Inmiddels heeft de Europese Commissie grote ambities geformuleerd voor het delen van data middels substantiële financiële middelen en een Data Governance Act en nog te presenteren Data Act.
782 Europese Commissie 2018.
783 Trommel, 20 april 2021: 28.

TomTom en is een aantal technologiestartups uitgegroeid tot grote bedrijven zoals Spotify en Zalando en uit Nederland Booking.com, Adyen en Takeaway. Er zijn in de EU echter geen grote gediversifieerde technologieplatformen zoals in de VS en China en voor veel Europese startups geldt dat zij met de Amerikaanse markt zijn verbonden door overnames of geïnvesteerd kapitaal.

Het Verenigd Koninkrijk, dat na de Brexit niet meer tot de EU behoort, heeft het meest ontwikkelde AI-ecosysteem van Europa. Het land is met name sterk in het doen van fundamenteel onderzoek, iets dat teruggaat tot het vroege onderzoek naar AI door wetenschappers als Alan Turing, naar wie ook het grote nationale AI-onderzoeksinstituut is vernoemd. Uit Brits onderzoek ontstond DeepMind, het geavanceerde AI-lab dat in 2014 door Google werd overgenomen. DeepMind is verantwoordelijk voor veel van de algoritmes die de laatste jaren stof hebben doen opwaaien, waaronder AlphaGo. Andere Europese landen met sterke AI-capaciteiten zijn Duitsland en Frankrijk. Duitsland is vooral sterk in het domein van robotica en gebruikt AI voor slimme toepassingen in fabrieken. Ook Frankrijk kent industriële toepassingen, maar zet vooral in op toepassingen in de zorg en defensie.

Een ander land dat internationaal een competitieve positie in AI heeft, is Japan. Ook daar is AI sterk gekoppeld aan de industriële sector en specifiek aan autofabrikanten. Canada is weer een ander land waar AI sterk is ontwikkeld. Net als in het Verenigd Koninkrijk ligt de basis daarvan in een sterk ecosysteem van fundamenteel onderzoek. Dat werd gedreven door de aanwezigheid van de prominente wetenschappers Geoffrey Hinton, Yann LeCun en Yoshua Bengio, wiens onderzoek door het Canadese instituut CIFAR gefinancierd werd.[784]

De AI-capaciteiten van Rusland zijn relatief beperkt, zeker wat betreft investeringen in onderzoek.[785] Tegelijkertijd heeft het land op specifieke domeinen binnen de AI wél een sterke positie. Zo kondigde het bedrijf United Instrument-Making Corporation in 2015 een groot onderzoeksproject aan op het gebied van AI en semantische data-analyse. De Russische Google, Yandex, gebruikt al jaren AI voor zoekresultaten. Het bedrijf ABBYY richt zich op de herkenning van tekstdata. VisionLabs is gespecialiseerd in gezichtsherkenning voor banken en de detailhandel. N-Tech.Lab heeft met het FaceN-algoritme in 2015 de eerste plaats behaald in een wereldwijde competitie in gezichtsherkenning.[786] Het programma van dit bedrijf koppelde de beelden van Russische burgers vanuit allerlei databronnen en platformen aan elkaar. Een conferentie in 2018 zette de

784 Uit een interview met Geoffrey Hinton in Ford 2018: 92.
785 Zie Mols 2019.
786 Bendett 2019: 171-172.

lijnen uit voor een Russische AI-strategie die zich richt op het ontwikkelen van expertise, training en onderwijs, het in kaart brengen van mondiale ontwikkelingen, en het gebruik van AI bij *war games*.[787]

Een belangrijke conclusie van dit korte overzicht is dat veel van de genoemde landen inzetten op wat wij in hoofdstuk 5 een nationale 'AI-identiteit' hebben genoemd. Inzetten op onderscheidende AI-capaciteiten zou ook de Nederlandse positie in AI kunnen versterken. Nederland zit niet in de groep van middelgrote landen als het Verenigd Koninkrijk, Duitsland, Frankrijk of Japan. Dat betekent niet dat het niet competitief is of kan zijn in het internationale speelveld. Ook relatief kleine landen kunnen zich sterke posities verwerven in AI, zeker wanneer zij zich op bepaalde domeinen specialiseren, en zo een AI-identiteit creëren. Als het aantal relevante AI-spelers afgezet wordt tegen de economische omvang, dan blijkt een aantal kleine landen zeer succesvol in AI te zijn. Opvallend zijn vooral Israël, Zuid-Korea en Singapore.[788]

Kernpunten – AI-capaciteiten

- AI is een complex fenomeen, maar we kunnen de capaciteiten van een land inschatten aan de hand van vijf variabelen: kwaliteit van fundamenteel onderzoek, beschikbaarheid van data, aanwezigheid van vereiste hardware, het ecosysteem voor bedrijven en een faciliterende/stimulerende overheid
- De VS en China zijn de twee wereldleiders die goed scoren op alle vijf de variabelen, maar duidingen over de relatieve positie van beide landen ten opzichte van elkaar lopen uiteen. De EU als geheel scoort ook goed, behalve als het gaat om een ecosysteem waarin bedrijven AI productief kunnen maken.
- Het Verenigd Koninkrijk, Duitsland, Frankrijk, Japan, Canada en Rusland zijn middelgrote spelers die op specifieke domeinen of toepassingen excelleren en daardoor een AI-identiteit hebben.
- Nederland behoort tot een categorie van kleinere landen die desondanks door hun specialismen ook relevante spelers op het wereldtoneel kunnen zijn.

787 Bendett 2019: 174.
788 Mols 2019: 16.

Nationale AI-strategieën

Het veld van AI is bijzonder dynamisch. Dat geldt niet alleen voor private spelers, maar juist ook voor overheden. Zoals we in hoofdstuk 2 zagen, heeft een groot aantal landen de laatste jaren AI-strategieën gepresenteerd, waaronder Nederland. In die strategieën komen weliswaar verschillende AI-gerelateerde zaken aan de orde, zoals AI-gebruik door de overheid en ethische principes, maar ze zijn er vaak primair op gericht de concurrentiekracht van een land te versterken.

We kunnen een aantal patronen uit de AI-strategieën destilleren. De Canadese onderzoeker Tim Dutton heeft verschillende van die documenten vergeleken en hij onderscheidt de volgende algemene thema's: onderzoek, talent, industriele strategie, ethiek, de toekomst van werk, data, AI-gebruik door de overheid en inclusie.[789] Daarvan zijn onderzoek, industriële strategie en talent de meest veelvoorkomende onderwerpen. Ook ethiek komt relatief vaak voor, maar de passages hierover zijn vaak vrij generiek. De invulling van ethiek en breder van publieke waarden lijkt iets dat met enige jaren vertraging volgt op de publicatie van de nationale strategieën.

Investeringen in onderzoek en talent komen in veel strategieën aan bod. De Duitse strategie kondigde bijvoorbeeld twaalf R&D-centra en honderd leerstoelen aan. De Amerikaanse universiteit MIT kondigde een investering aan van een miljard dollar in een *AI College*. Coördinatie en samenwerking in onderzoek komen ook veel terug. De Franse strategie voorziet in vier interdisciplinaire AI-instituten en in Canada werkt CIFAR samen met verschillende instituten om onderzoek te coördineren. In 2015 werd in het Verenigd Koninkrijk het Alan Turinginstituut opgericht, waar een groeiend aantal onderzoekscentra aan is verbonden.

Een ander patroon dat de verschillende documenten laten zien, is dat ze AI in een breder kader van technologische ontwikkeling plaatsen. De Chinese strategie hangt bijvoorbeeld samen met andere plannen voor sleuteltechnologieën, zoals *Made in China 2025*. De Japanse strategie plaatst AI binnen de 'vierde Industriële Revolutie'. Hetzelfde geldt voor de Zuid-Koreaanse strategie, die ook spreekt van een 'Intelligent Information Society'. Het SAPAI is inmiddels opgenomen in de bredere digitaliseringsstrategie van de Nederlandse overheid.

In lijn met het idee van een AI-identiteit valt ook op dat veel landen de ontwikkeling van AI in hun strategie koppelen aan sectoren en domeinen waar het land reeds competitief in is. Duitsland's plan *Strategie Künstliche Intelligenz*

789 Dutton, 28 juni 2018.

der Bundesregierung legt bijvoorbeeld de nadruk op de implementatie van AI in de zware industrie. Het volgt op eerdere strategische initiatieven van het land, zoals *Industrie 4.0,* dat gericht was op robotica en slimme fabricatie. Als grote producent van machines, infrastructuur en transporttechnologie wil Duitsland leidend zijn als het erom gaat die sectoren slim te maken.

De Japanse strategie voor AI legt de nadruk op drie domeinen, waaronder mobiliteit. Met bedrijven als Toyota, Nissan, Honda en Mitsubishi heeft Japan veel te winnen bij de implementatie van AI in die markt. Hetzelfde geldt voor Frankrijk, met grote autofabrikanten als Peugeot, Renault en Citroën. In het rapport dat de wiskundige Cédric Villani schreef voor de Franse regering, licht hij vier gebieden uit voor Frankrijk. Mobiliteit is daar één van, maar ook defensie, een andere sector waar de Franse economie sterk in is. Een derde is gezondheid, waartoe onder meer een zorgdata-Hub wordt opgezet waarin gegevens van zorgverleners, ziekenhuizen, zorgverzekeraars, farmaceutische bedrijven, laboratoria en andere relevante partijen bijeen worden gebracht.[790] Ook hier wordt voortgeborduurd op nationale krachtbronnen. De Franse economie wordt gekenmerkt door een hoge mate van centralisatie (*dirigisme*) en dat betekent op het gebied van zorg dat het land enorme gecentraliseerde databases heeft en kan ontwikkelen, die als basis kunnen dienen voor AI. Dat is in veel andere landen niet het geval.

Ook andere landen die evenals Frankrijk sterk zijn in defensie, richten zich op die sector. Israël heeft officieel nog geen AI-strategie, maar beschikt wel over sterke capaciteiten en heeft de ambitie uitgesproken om leidend te worden in AI op het gebied van defensie en cybersecurity. Zoals we zagen, is het organiseren van AI-oorlogsspellen onderdeel van Ruslands strategie. Landen als Rusland en China hebben bovendien een overheid die veel grip heeft op de eigen bevolking. Niet voor niets zijn beide landen dan ook erg sterk in AI-toepassingen op het gebied van gezichtsherkenning. Daar komen wij in paragraaf 8.2 op terug.

Een ander patroon in veel AI-strategieën is om juist de nadruk te leggen op domeinen waar grote maatschappelijke vraagstukken spelen en waar AI een antwoord op kan bieden. Dit lijkt de reden waarom de pijler van zorg in de Japanse strategie is opgenomen. Japan is namelijk de meest vergrijsde samenleving ter wereld en kampt daardoor met een toenemende zorgbehoefte. AI zou kunnen helpen om daarin te voorzien. De Indiase strategie heet *AI for All* en heeft inclusie als expliciet doel. Dat is een grote uitdaging voor dit land waarin een grote economische en sociale ongelijkheid heerst. Allerlei recente digitale strategieën van het land, zoals een programma voor financiële inclusie en

790 France Strategie 2019.

dienstverlening middels biometrische data, zijn er dan ook op gericht om een inclusievere samenleving te realiseren. In het kader van die ambitie moeten we de Indiase strategie voor AI bezien. Naast de drie eerder genoemde domeinen, is ecologie een vierde domein in de Franse AI-strategie. Ook daarin is de Franse economie sterk vertegenwoordigd, vooral op het gebied van energie. Daarnaast is ecologie sinds het Parijsakkoord ook een mondiale uitdaging waar Frankrijk zich graag op profileert.

Een laatste patroon in verschillende AI-strategieën is beleid voor het toepassen van onderzoek in praktische en commerciële contexten. Een van de vijf pijlers van de Britse *AI Sector Deal* is het opzetten van een AI Council die als taak heeft de samenwerking tussen universiteiten en het bedrijfsleven te verbeteren. Een andere aanpak is het Canadese SCALE.AI, dat onderdeel is van het nationale beleid van 'superclusters'. Bedrijven in de detailhandel, de maakindustrie, transport, infrastructuur en ICT worden hierin bij elkaar gebracht om middels AI en robotica slimme logistieke ketens te ontwikkelen die het Canadese bedrijfsleven competitiever maken. Een ander innovatief project is Singapores *100 Experiments*. Daarin worden bedrijven uitgenodigd om vraagstukken aan te leveren waarbij AI uitkomst kan bieden, maar waar op dit moment nog geen algemeen toegankelijke producten voor zijn. Een product zou wel binnen negen tot achttien maanden gebouwd moeten kunnen worden. Bedrijven worden vervolgens gekoppeld aan Singaporese AI-onderzoekers met ondersteunende financiering van de overheid.

De Nederlandse aanpak in het SAPAI sluit op verschillende punten aan bij deze mondiale trends. Het document bestaat uit drie delen. Het eerste deel gaat over het grijpen van de economische kansen die AI biedt. Daarbij horen investeringen. Het tweede deel gaat over de voorwaarden die nodig zijn om AI te laten floreren in Nederland, waaronder het stimuleren van talent. Het laatste deel gaat over de maatschappelijke fundamenten waar AI betrekking op heeft: de publieke waarden die geborgd moeten worden of die met AI versterkt kunnen worden.

Hoe verhoudt het SAPAI zich tot andere nationale strategieën? Net als in veel andere documenten zijn onderzoek en talent centrale pijlers. Door de koppeling met de Nederlandse AI Coalitie is er in de strategie ook veel aandacht voor samenwerkingsverbanden tussen onderzoek, overheid en bedrijfsleven. Minder dan andere strategieën kiest het SAPAI voor aandachtspunten of specifieke sectoren, wat wij een AI-identiteit noemen. Ook de thematiek van ethiek en publieke waarden wordt in het SAPAI behandeld en daarvoor worden verschillende vervolgtrajecten benoemd. De opname van het SAPAI in de digitaliseringsagenda van de overheid past ten slotte bij de mondiale trend om AI in te bedden in een bredere visie op nieuwe technologieën.

Kernpunten – Nationale AI-strategieën
- Sinds 2017 zijn er meer dan zestig nationale AI-strategieën gepubliceerd.
- In die documenten is een aantal patronen te onderscheiden: een nadruk op investeringen in onderzoek en talent, het plaatsen van AI in een breder kader van technologische ontwikkelingen, de koppeling aan sectoren waar een land al competitief in is, identificatie van uitdagingen waarvoor AI gebruikt kan worden en de commercialisering en toepassing van onderzoek.
- Het Nederlandse Strategisch Actieplan voor AI (SAPAI) is in oktober 2019 gepubliceerd, waarna de overheid verschillende AI-trajecten in gang heeft gezet.

Een internationale AI-race?

Een prominent onderdeel van veel strategieën is de internationale positionering van het betreffende land. Omwille van de nationale competitiviteit lijkt er zelfs een race tussen landen gaande om leider te worden in AI. Een daarop ingestoken zinsnede komt dan ook in veel nationale strategieën voor. De Chinese strategie spreekt van het ontwikkelen van een "*first-mover advantage* in de ontwikkeling van AI" en de Amerikaanse strategie heeft het over het "accelereren van het Amerikaanse leiderschap in AI".[791] Ook veel auteurs gebruiken dit frame van een mondiale race.[792] Kai Fu Lee ziet een analogie met de ruimtevaart tijdens de Koude Oorlog. De overwinning in Go op Lee Sedol kunnen we zien als China's Spoetnikmoment en de presentatie van de nationale AI-strategie een aantal maanden later was het equivalent van de speech waarin de toenmalige president Kennedy Amerika opriep om een mens op de maan te zetten.[793] Het Spoetnikmoment leidde tot de oprichting van NASA en DARPA, de innovatietak van het Amerikaanse leger.[794] Net als ruimtevaart toen is AI nu de focus van veel toekomstige innovatie.

Verschillende ontwikkelingen zijn inderdaad te begrijpen als een race. Landen beconcurreren elkaar in de toegang tot talent. Het Duitse beleid om bijvoorbeeld meer leerstoelen op het gebied van AI op te zetten, kan nadelig zijn voor het behoud van getalenteerde onderzoekers in Nederland. In de media is er uitgebreid gedebatteerd over het vertrek van Nederlandse onderzoekers naar het buitenland.[795] Een studie van het Rathenau Instituut laat overigens zien dat er

791 Smuha 2019: 2.
792 Zie bijvoorbeeld Walch 2018; Harari 2019.
793 Lee 2018: 98.
794 Weinberger 2019.
795 Hueck 2 september 2018; De Rijke 8 april 2019.

nog geen sprake is van een netto-uitstroom van onderzoekers uit Nederland.[796] De verdere uitvoering van AI-strategieën kan hier in de toekomst wel druk op zetten en het is onderwerp van debat of de aangekondigde Nederlandse investeringen op dit punt toereikend zijn.[797] De Cyber Security Raad constateert dat veel Nederlandse AI-opleidingen een maximum aantal studenten hebben en roept op om te investeren in het opleiden van talent.[798]

Ook het bredere beleid gericht op competitiviteit kan als een race begrepen worden. Een late omarming van AI kan namelijk leiden tot verlies van verdienvermogen. De AWTI heeft de Nederlandse overheid eerder opgeroepen om gerichter te investeren in sleuteltechnologieën.[799] Het is bovendien mogelijk dat een achterstand in de loop der tijd versterkt wordt door bijvoorbeeld netwerkeffecten en afhankelijkheden. Specifiek voor AI geldt dat toegang tot grote hoeveelheden goede data leidt tot betere algoritmes en dat zo een vicieuze cirkel ontstaat die voor andere partijen moeilijk te doorbreken is. Hiernaar wordt vaak verwezen als de *winner-takes-all*-dynamiek van AI.

Of denk aan de invloed van AI op specifieke sectoren. Als Amerikaanse bedrijven bijvoorbeeld door investeringen in AI succes boeken met zelfrijdende auto's, kan het zeer nadelig zijn voor de Duitse economie als de grote eigen lokale auto-industrie daarbij achterblijft. De focus die wij hiervoor tegenkwamen in AI-strategieën op sectoren waar landen al goed in zijn, vertrekt dus zowel vanuit de kansen als de risico's: landen met grote autofabrikanten hebben expertise en data die ze kansen biedt voor het ontwikkelen van zelfrijdende auto's. Pakken ze die kansen niet, dan kan hun sector op achterstand komen en marktaandeel verliezen. Hetzelfde geldt voor de Nederlandse landbouwsector. China zet in op innovatieve technologie in de landbouw en Amerikaanse bedrijven als John Deere hebben door hun machines toegang tot data over de Nederlandse land- en tuinbouw. Er ligt voor Nederland dus een kans op het gebied van AI in de land- en tuinbouw, en tegelijkertijd is er de dreiging dat onze traditioneel sterke positie in deze sector verzwakt.

Ten slotte is de metafoor van een mondiale race mogelijk ook adequaat voor het militaire domein, omdat AI een bepaald land een strategisch voordeel kan bieden op andere landen. In paragraaf 8.2 gaan we hier verder op in. De metafoor kent echter ook serieuze tekortkomingen. Er kleven namelijk verkeerde

796 Rathenau Instituut 2021.
797 Kees Verhoeven in: Hueck en Van Wijnen 8 oktober 2019.
798 Cyber Security Raad 2019.
799 AWTI 2020.

veronderstellingen aan over de wereldwijde ontwikkeling van AI. Figuur 8.2 geeft deze problemen weer; we bespreken ze hieronder.

Figuur 8.2 Vier problemen met het frame van een AI-race

Een eerste tekortkoming van het beeld van een race is dat dit suggereert dat er maar een enkele winnaar kan zijn. De ontwikkeling van AI wordt voorgesteld als een *zero sum*-situatie. Zoiets kan het geval zijn bij specifieke doelen, bijvoorbeeld om als eerste de maan te bereiken, maar is minder toepasselijk als het gaat om de ontwikkeling van een systeemtechnologie. Ook Kai-Fu Lee, die zoals we zagen de analogie met ruimtevaart maakt, suggereert dat een race niet de juiste benadering is. De wereld van AI lijkt volgens hem uiteindelijk meer op de Industriële Revolutie of op de opkomst van elektriciteit dan op de *'space race'* tijdens de Koude Oorlog. Net zoals bij elektriciteit is er niet zomaar sprake van *zero sum*. De winst van de één is niet per se het verlies van de ander. Technologieën verspreiden zich over de hele wereld en leiden op allerlei plekken tot vooruitgang en vergroting van de welvaart.[800] We zouden dit zelfs kunnen beweren van de *space race*. Ook al kon er maar één land als eerste op de maan landen, de innovaties die ruimtevaart mogelijk maakten, zoals satellieten en GPS-technologie, leveren profijt op voor mensen overal ter wereld.

Bij deze voorbeelden waren er wel degelijk bepaalde landen die meer voordeel hadden dan andere. Bijvoorbeeld omdat zij de innovatieve bedrijven hadden die industriële producten, energiebronnen of ruimtevaarttechnologie naar andere landen exporteerden. Het is echter belangrijk om te benadrukken dat de voordelen van die technologieën ook door burgers elders werden gedeeld. Dit

800 Lee 2018: 227.

perspectief verlegt de aandacht van de ontwikkeling naar de diffusie van een technologie.

Bij AI gaat de aandacht vaak uit naar competities op het meest geavanceerde niveau, waar slechts de leidende bedrijven van de rijkste landen aan meedoen. Maar het daadwerkelijke effect van AI op de wereld vindt in hoge mate plaats door veel minder geavanceerde vormen die bovendien veel wijder verspreid zijn. Als we de analogie van elektriciteit volgen, is er misschien maar een kleine groep landen die nucleaire energie kan ontwikkelen, want die vorm van energieopwekking is uiterst geavanceerd, risicovol en vergt grote investeringen. Een groot palet aan manieren om elektriciteit op te wekken is echter breed beschikbaar en dat biedt voordelen aan burgers en markten voor bedrijven overal ter wereld. Oftewel, innovatie is vaak geconcentreerd op een paar plekken, de productie ervan is al veel meer gedistribueerd en het gebruik ervan nog veel wijder.[801]

Een focus op diffusie is ook belangrijk omdat dit onderwerp andere vragen met zich meebrengt dan de focus op de technologische 'frontlinie' (*frontier*) van AI, de plek waar innovatie plaatsvindt. Het ontwikkelen van wereldwijd leidende laboratoria is iets anders dan zorgen dat AI breed in de samenleving wordt ingebed. Voor de Amerikaanse economie geldt bijvoorbeeld dat die op veel domeinen technologisch aan de frontlinie staat, maar dat de brede bevolking daar minder van profiteert dan in andere landen.[802] En Nederland is historisch gezien op veel gebieden niet de ontwikkelaar van een nieuwe technologie geweest, maar wel erg succesvol in de toepassing en verspreiding ervan.[803]

Een andere implicatie van de focus op diffusie is dat technische expertise niet meer alleen gaat over die van uitvinders en grote laboratoria, maar vooral ook van belang is op het gebied van onderhoud en herstel. Een zeer groot gedeelte van de mensen die werken met elektriciteit of IT, doen dat als reparateurs of onderhouders van systemen. Ook bij AI is aandacht nodig voor deze belangrijke vorm van technische expertise.

Als we redeneren in termen van diffusie in plaats van de technologische frontlinie, blijkt ook de reële zorg dat bepaalde groepen in de maatschappij niet in deze ontwikkeling kunnen meekomen. Het idee van een mondiale race richt onze aandacht op een strijd tussen landen en vergelijkt hun leidende bedrijven,

801 Edgerton 2008: 80.
802 Hall en Soskice 2001.
803 WRR 2013.

zonder oog te hebben voor hun bredere bevolking. Het kan daarmee het vraagstuk van ongelijkheid binnen landen ten aanzien van AI overschaduwen.

Een tweede probleem met het beeld van een mondiale race is dat het suggereert dat iedereen op weg is naar hetzelfde doel. In de eerste plaats is het onduidelijk wat dat doel zou moeten zijn. Zoals duidelijk is geworden in dit rapport, is AI een complex fenomeen met enorm veel mogelijke toepassingsgebieden. Succes is dan ook heel moeilijk te definiëren. Leiderschap in AI is veel lastiger te identificeren dan de eerste maanlanding. Succes in AI kan langs verschillende assen plaatsvinden en dat maakt een beeld met een enkele eindstreep problematisch.

Landen kunnen bovendien ook heel verschillende trajecten volgen. In hoofdstuk 3 zagen we dat de ontwikkeling van elektriciteit in continentaal Europa voor andere toepassingen plaatsvond en met andere organisatiemodellen dan in de VS. Ook bij AI-strategieën positioneren landen als Canada en het VK zich bijvoorbeeld op het domein van fundamenteel onderzoek, terwijl andere landen zich meer op specifieke sectoren of toepassingsgebieden richten. Het idee van een AI-identiteit waarbij landen zich van elkaar onderscheiden in hun AI-capaciteiten, verdraagt zich niet met het beeld van een AI-race. Dat beeld suggereert dat iedereen op hetzelfde pad zit en dat maakt ons blind voor de diverse manieren waarop met AI succes geboekt kan worden en nationale competitiviteit kan worden versterkt.

Belangrijker wellicht dan de voorgaande bezwaren is dat het beeld van een mondiale race een tegenstelling suggereert tussen competitiviteit en het borgen van publieke waarden. Vanuit het idee dat sommige landen dominant dreigen te worden, wordt vaak beweerd dat landen die achterlopen vaart zouden moeten maken en zich minder moeten laten 'ophouden' door discussies over het beschermen van rechten. Dat zouden zij zich niet kunnen permitteren omdat hun achterstand dan groeit, zo is de stelling. Denk aan het beperken van de toegang tot data om privacyoverwegingen of het aan banden leggen van experimenteren met surveillance om de vrijheid van burgers te garanderen. Nick Bostrom benadrukt dat de racedynamiek ten koste kan gaan van voorzichtigheid en veiligheid.[804]

Alhoewel dergelijke spanningen in de praktijk zeker voorkomen, is het gevaarlijk en onterecht om het versterken van competitiviteit en het beschermen

804 "This is one of the concerns with a racing dynamic, where you have a lot of different competitors racing to get to some kind of finish line first – in a tight race you are forced to throw caution to the wind. The race would go to whoever squanders the least effort on safety, and that would be a very undesirable situation" (Ford 2018: 113).

van burgerrechten tegenover elkaar te plaatsen. In de eerste plaats is het niet duidelijk of economische competitiviteit die burgerrechten schendt, op de lange termijn houdbaar is. Extreme surveillance die criminaliteit reduceert tegen de prijs van individuele vrijheid, is de moeite niet waard. Zeker in landen als Nederland, met sterke democratische tradities, zal dergelijke innovatie verzet opleveren, hoe effectief deze ook mag zijn. Daarnaast moet benadrukt worden dat het economisch succes van een innovatie juist gestimuleerd kan worden door publieke waarden te borgen. Dat zagen we ook bij eerdere systeemtechnologieën. Alhoewel er aanvankelijk bijvoorbeeld weerstand was tegen veiligheidsmaatregelen in auto's – deze zouden de prijzen opdrijven en innovatie tegengaan – leidden die er wel toe dat burgers meer vertrouwen kregen in auto's, waardoor zij er meer gebruik van gingen maken.

Precies dit argument gebruikt de Europese Unie ook in haar meer ethische-benadering van AI. De Europese Commissie stelt: "*Building on its reputation for safe and high-quality products, Europe's ethical approach to AI strengthens citizens' trust in the digital development and aims at building a competitive advantage for European AI companies.*" En Pekka Ala-Pietilä, voorzitter van de High-Level Expert Group on AI van de Commissie, stelt: "*Ethics and competitiveness go hand in hand. Businesses cannot be run sustainably without trust, and there can be no trust without ethics. And when there is no trust, there is no buy-in of the technology, or enjoyment of the benefits that it can bring.*"[805] Later in dit hoofdstuk gaan we nader in op de rol van de Europese Unie bij wetgeving en ethiek. Los van enkele punten van kritiek kunnen we hier wel concluderen dat de Europese Commissie correct constateert dat competitiviteit en publieke waarden niet noodzakelijk op gespannen voet met elkaar staan.

De aanpak van de EU wordt weleens bekritiseerd, omdat deze de ethische aspecten meer benadrukt dan competitiviteit in AI te ontwikkelen. Aan dat laatste werkt de EU en er wordt in het openbaar bediscussieerd of dat in voldoende mate gebeurt. Andere terechte kritiek op dit beleid is dat de EU zich iets te gemakkelijk positioneert als de ethische leider in de wereld tussen de marktgedreven VS en het staatsgedreven China. Daarmee veronachtzaamt ze de acties van andere landen op dit gebied. Landen als Japan, Canada, Dubai, Singapore en Australië hebben bijvoorbeeld ook verschillende richtlijnen voor AI ontwikkeld. Zelfs China, dat vaak wordt neergezet als een land waar ethische normen laag op de agenda staan, heeft in 2019 de *Beijing Principes for Ethical AI* gepubliceerd.[806]

805 Smuha 2019: 15.
806 Smuha 2019: 19.

Verbonden met het idee dat een race een enkele winnaar heeft, is de gedachte dat dominante landen de ontwikkeling van een technologie binnen hun grenzen kunnen afbakenen. We hebben bij eerdere systeemtechnologieën gezien dat dit nooit is gelukt en dat de ontwikkeling van een revolutionaire technologie altijd een mondiale aangelegenheid is geweest, gedreven door onderzoekers en bedrijven uit verschillende landen. De pogingen van overheden om innovatie te nationaliseren slaagden niet en werkten vaak zelfs averechts: de bedrijven van dat land raakten hun marktpositie kwijt vanwege exportcontroles en het nationalistische overheidsbeleid moedigde concurrentie aan.

Deze dynamiek lijkt nu ook plaats te vinden in het Chinees-Amerikaanse handelsconflict waar AI een onderdeel van uitmaakt. De Amerikaanse regering poogt China's opkomst in AI te remmen of tegen te houden door met name chipsbedrijven als Intel te verbieden om hun geavanceerde producten aan China te verkopen. Er is ook Amerikaanse druk op Nederlandse bedrijven als ASML om de export naar China te beperken. Het is echter heel goed mogelijk dat dit de Chinese ambities om een eigen chipssector te ontwikkelen zal versnellen. Het Amerikaanse beleid kan bovendien partnerschappen met derde landen in Europa of Azië versterken. Het handelsconflict noopt Europese landen ertoe zich in dit mondiale speelveld te positioneren.

Kernpunten – Internationale AI-race

- Een aantal ontwikkelingen op het internationale toneel, zoals de toegang tot talent en defensie, kunnen begrepen worden als een race. Tegelijkertijd kent dit beeld serieuze tekortkomingen.
- Het beeld van een race suggereert onterecht een zero sum-situatie en negeert het belang van diffusie.
- Ook zijn niet alle landen op weg naar hetzelfde doel en is er dus niet één 'winnaar'.
- De metafoor van een race suggereert bovendien onterecht dat er een spanning is tussen concurrentiekracht en het zorgdragen voor publieke waarden.
- Ten slotte is het onmogelijk om AI-innovatie succesvol binnen de eigen grenzen te laten plaatsvinden.

Van competitie naar samenwerking

De historische les is in ieder geval dat geen land in staat zal zijn om AI volledig te binnen de eigen grenzen af te bakenen en verder te ontwikkelen, zelfs niet de VS of China met hun geavanceerde ecosystemen. Voor landen als Nederland is het des te meer uitgesloten dat AI in afzondering ontwikkeld kan worden. Als klein land zijn wij gebaat bij openheid en internationale samenwerking. Sterker

nog, een focus op internationale samenwerking, in het bijzonder in Europees verband, kan de concurrentiepositie van een land als Nederland juist versterken. Nederland zou daar dan ook meer op kunnen inzetten om zich internationaal uitdrukkelijker te profileren. Dat betekent een geïntegreerd beleid van 'AI-diplomatie'. We onderscheiden vijf domeinen waarop dat beleid plaats kan vinden: fundamenteel onderzoek, commerciële toepassingen, regulering, ethische richtlijnen en standaarden.

Tekstbox 8.1 – CLAIRE

CLAIRE is een samenwerkingsverband van AI-wetenschappers opgericht door Holger Hoos, hoogleraar aan de Universiteit Leiden, Philipp Slusallek (Duitsland) en Morten Irgens (Noorwegen). Zij schreven een visiedocument dat door meer dan 550 AI-experts is ondertekend. De doelen van CLAIRE zijn het ontwikkelen van excellentie in alle domeinen van AI door heel Europa met een mensgerichte focus. Een sterke Europese onderzoeksorganisatie is volgens de oprichters van belang voor de ontwikkeling van AI in Europa; een achterstand op het gebied van AI zou kunnen leiden tot negatieve economische gevolgen, een academische braindrain, minder transparantie en toenemende afhankelijkheid van buitenlandse technologieën.

CLAIRE moet die doelen bereiken door een netwerk op te zetten van excellentiecentra in heel Europa, regionale centra met eventuele specialisaties en een centrale hub of Europees '*lighthouse*'. Op die manier moet het project voor AI zijn wat CERN is voor de deeltjesfysica: een centraal instituut waar de state-of-the-artinfrastructuur voor AI aanwezig is.

In 2020 werd het hoofdkantoor gevestigd in Den Haag. Dit hoofdkantoor ondersteunt de activiteiten van CLAIRE in de rest van Europa en coördineert de andere kantoren. Daarnaast richt het hoofdkantoor zich vooral op AI in de publieke sector, AI-computing en data-infrastructuur, en de ontwikkeling van AI voor sociaal welzijn. Via een subsidie van het ministerie van Binnenlandse Zaken draagt de Nederlandse overheid bij aan dit Europese onderzoeksverband.

Samenwerking op het gebied van fundamenteel onderzoek is het meest duidelijk. Door internationale samenwerking te stimuleren en naar zich toe te trekken, versterkt een land op directe wijze de eigen ontwikkeling van AI. Op dit gebied participeert ons land al aan een aantal internationale projecten. De twee prominentste projecten in Europa zijn ELLIS en CLAIRE. Dat zijn samenwerkingsverbanden voor breed AI-onderzoek en specifiek voor machine learning.

CLAIRE wordt beschreven als 'het CERN voor AI' (zie tekstbox 8.1). Nederland doet het op dit gebied goed: in 2020 werd het hoofdkantoor van CLAIRE in Den Haag gevestigd.

Een tweede domein waarbij samenwerking het verdienvermogen kan versterken, zijn commerciële projecten. Te denken valt aan het opzetten van nieuwe diensten en organisaties in internationale samenwerkingsverbanden, maar ook aan de versterking en coördinatie rondom bestaande bedrijven. Binnen de EU gebeurt er de laatste tijd veel op dit terrein, vanuit het streven naar digitale strategische autonomie.[807] Onderdeel van die ambitie is het project Gaia-X rondom cloud- en data-infrastructuur (zie tekstbox 8.2). Er zijn vergelijkbare Europese projecten rondom cybersecurity.[808]

Tekstbox 8.2 – Gaia-X

Gaia-X is een project gestart in Duitsland, waar Frankrijk zich bij heeft gevoegd en meer Europese landen zich bij aansluiten. Het startdocument werd in oktober 2019 gepresenteerd en op 15 september 2020 heeft een groep van 22 organisaties een *incorporation paper* ondertekend, waaronder Duitse bedrijven als Bosch en Siemens en Franse bedrijven als Orange en Atos. Details over het project moeten nog duidelijker worden, maar de doelen ervan zijn het versterken van de Europese data-soevereiniteit, het verminderen van de buitenlandse afhankelijkheid (door lock-in), het aantrekkelijker aanbieden van cloudservices en het creëren van een ecosysteem voor innovatie. Hiertoe voorziet het project in een infrastructuur die niet een zozeer zelf een alternatief moet bieden voor Amerikaanse clouddiensten, als wel het voor Europese partijen gemakkelijker moet maken om daarmee te concurreren, waardoor Europese data vervolgens in eigen beheer blijven. De infrastructuur levert gemeenschappelijke regels, normen en technologie. Binnen het project worden verschillende toepassingsdomeinen geïdentificeerd zoals *Sustainable Finance* en *Ambient Assisted Living*; bij beide zal AI worden ingezet. Een jaar na de presentatie van het startdocument, ondertekende de Nederlandse regering op 15 oktober 2020 een verklaring waarin zij aangeeft de ontwikkeling van een nieuwe Europese cloudinfrastructuur te steunen.

807 Sheikh en Timmers 2020.
808 Voor een vertaling daarvan naar de Nederlandse context zie: Timmers en Dezeure 2021, in opdracht van de Cyber Security Raad.

Bij commerciële projecten valt ook te denken aan meer samenwerking rondom bestaande bedrijven die actief zijn op het gebied van AI of daarmee verbonden zijn, zoals de Scandinavische bedrijven Nokia en Ericsson die telecominfrastructuur leveren. In Nederland kan gedacht worden aan chipsbedrijven als ASML en NXP. Vanuit het idee van digitale strategische autonomie kan Nederland bijdragen aan het beschermen en versterken van Europese bedrijven in dit domein en mogelijk zelfs verdere Europese samenwerking voorstaan. Alhoewel er terechte twijfel is bij het voeren van industriepolitiek, heeft Europa in het verleden laten zien dat dit succesvol kan zijn. Vanuit een positie van achterstand hebben Europese landen door samen te werken de luchtvaartgigant Airbus en het satellietnavigatiesysteem Galileo opgezet.

Een derde internationaal domein waar een land zich op kan richten in de ambitie om zich middels samenwerking te positioneren, is wetgeving. Dit is bij uitstek een domein waar de Europese Unie meerwaarde kan realiseren en over bevoegdheden beschikt. *Regulatory power* wordt vaak gezien als een vorm van soft power die het blok tot beschikking heeft.[809] Een relevant recent voorbeeld is natuurlijk de AVG, die wereldwijd een standaard heeft gezet voor gegevensbescherming. Op deze Verordening inzake de verwerking van persoonsgegevens is echter ook de nodige kritiek geuit. Volgens sommigen plaatst deze de EU op een grotere achterstand ten opzichte van China en de VS omdat ze innovatie met behulp van persoonsgegevens bemoeilijkt.[810] Wat echter duidelijk is, is dat de EU hiermee wereldwijd normen heeft bepaald en een discussie heeft aangewakkerd.[811] Verschillende landen, maar ook Amerikaanse staten hebben de onderliggende principes uit de AVG overgenomen en de Verordening heeft zelfs het Chinese databeschermingsbeleid beïnvloed.

Op de bijeenkomst van de G20 in 2019 in Japan stelde Angela Merkel dat het de taak van de volgende Europese Commissie zal zijn om, in opvolging van de AVG, wetgeving op te stellen voor "vertrouwenswaardige AI" (*trustworthy AI*).[812] Onderzoeker Nathalie Smuha spreekt in dit verband ook van "*regulatory*

809 Jan Zielonka spreekt in deze context zelfs over de EU als een "*regulatory empire*". (Zielonka 2008: 474) Anu Bradford noemt de wereldwijde invloed van de EU "The Brussels effect". Zonder internationale instituties te gebruiken of internationaal de samenwerking te zoeken kan de EU volgens haar regulering afkondigen, die ingebed raakt in de wettelijke kaders voor markten wereldwijd. Het effect is een 'europeanisering' van veel aspecten van de wereldwijde handel (Bradford 2020).

810 Smuha 2019.

811 Lee Bygrave legt precies uit hoe dit gewerkt heeft bij de GDPR, daarbij refererend aan het werk van Bradford. Bygrave legt daarnaast sterk het accent op de Europese denkkracht en de rol van de Raad van Europa in het geheel (Bygrave 2020).

812 Smuha 2019: 17.

co-opetition", een internationaal proces van samenwerking en competitie op het gebied van de regulering van AI, waaronder wetgeving.[813]

Een dergelijk proces vindt niet alleen plaats met betrekking tot wetgeving, maar ook rondom een vierde domein van internationale samenwerking: richtlijnen[814] en (ethische) principes. Op dit gebied gebeurt ontzettend veel, maar een paar prominente trajecten springen in het oog. Zo nam de OESO in mei 2019 een set gemeenschappelijke ethische principes voor AI aan. Het was daarmee de eerste set intergouvernementele beleidsrichtlijnen op het gebied van AI. Een maand later werden deze richtlijnen formeel omarmd door de G20. Ook is er een traject van UNESCO rondom een mondiale ethische code voor AI. De Raad van Europa ten slotte heeft een ad-hoccomité voor AI (CAHAI) opgesteld dat uiteindelijk tot doel heeft om in samenwerking met alle lidstaten tot bindende regels te komen voor de bescherming van mensenrechten, democratie en de rechtsstaat.[815]

Omdat het laatste domein van internationale samenwerking doorgaans minder aandacht krijgt, gaan wij er hier uitgebreider op in. Uit onze analyse van eerdere systeemtechnologieën werd namelijk het belang duidelijk van expertfora gericht op (technische) standaarden. Meer op de achtergrond is er enorm veel dynamiek in de fora waar standaardisering voor AI wordt ontwikkeld. In het vorige hoofdstuk keken we naar standaarden vanuit de rol die ze spelen bij de regulering van een nieuwe technologie. Standaarden zijn bijvoorbeeld een manier om de interoperabiliteit van een technologie te versterken. Hier kijken we niet naar de functies ervan, maar naar de dimensie van internationale samenwerking en de invloed ervan op de positionering van een land. Het vermogen om standaarden te bepalen heeft namelijk ook grote invloed op de competitiviteit van een land. De eigen bedrijven die die standaarden al volgen, krijgen zo een *first-mover advantage*. In plaats van elders ontwikkelde standaarden te implementeren, geeft het landen grip op de standaarden die in eigen land gelden. Bovendien levert het invloed op over andere landen die die standaarden moeten volgen en daardoor ontstaan mogelijk lock-ineffecten.

Zoals we in deel 1 van dit rapport zagen, is er bij de historische ontwikkeling van systeemtechnologieën een continue spanning geweest over de invulling van standaarden tussen technocratisch georiënteerde experts aan de ene kant en nationale politici en ambtenaren aan de andere kant. Hoe ziet het huidige internationale speelveld rondom standaarden eruit?

---

813 Smuha 2019: 26.

814 We doelen hier niet op de Richtlijn als instrument van EU-wetgevingsbeleid, maar als set van minder formele regels dan wettelijke bepalingen.

815 Smuha 2019: 21.

Opvallend is dat Europa daar mondiaal een leidende rol speelt. Het heeft zelfs '*standard power*'.[816] Dit heeft te maken met de lange historische ontwikkeling op dit gebied en het daarin ontwikkelde proces voor de integratie van standaarden over landsgrenzen heen, aanvankelijk binnen Europa, maar nu dus ook internationaal. Wat hier ook aan bijdraagt, is het specifieke Europese model van standaardisering. Dat kenmerkt zich door publiek-private samenwerking. Standaarden worden ontwikkeld door private standaardisatieorganisaties die daarvoor licenties krijgen van overheden. Elk land heeft een afzonderlijke organisatie ingesteld voor een specifiek doel. Voor algemene standaarden is dat in Nederland bijvoorbeeld de NEN (Nederlands Normalisatie Instituut) en in Duitsland de DIN (Deutsche Institut für Normung). Daarnaast zijn er standaardisatieorganen specifiek voor een sector, zoals het Duitse DKE voor de elektrotechniek (Deutsche Kommission Elektrotechnik Elektronik Informationstechnik).

Vervolgens is er in Europa sprake van een duidelijke hiërarchische structuur. Boven de nationale lichamen staan Europese organisaties: boven de NEN staat de CEN (Comité Européen de Normalisation) en boven DKE staat CENELEC (European Committee for Electrotechnical Standardization). Daarboven staan weer mondiale organisaties, respectievelijk de ISO (International Organization for Standardization) en de IEC (International Electrotechnical Commission). Middels internationale overeenkomsten worden deze niveaus gestroomlijnd en heeft de hogere laag voorrang bij het bepalen van standaarden.[817] De organen daaronder geven steeds nadere invulling aan die eisen.

Het Europese model verschilt hierin sterk van het Amerikaanse model, dat een veel sterkere private grondslag heeft. Daardoor is er in een bepaald domein vaak sprake van een veelheid aan standaardisatielichamen die met elkaar concurreren. Het gebrek aan coördinatie maakt het systeem wereldwijd minder invloedrijk dan het Europese systeem. Standaarden voor AI worden primair ontwikkeld in drie fora: een gezamenlijk initiatief van ISO en IEC, de ingenieursorganisatie IEEE Standards Association en de ITU, de telecommunicatieorganisatie van de VN.[818]

Een aantal zaken vlat op aan het internationale speelveld van standaardisering rondom AI. Het eerste is dat China een steeds grotere rol in dit veld speelt. Invloed uitoefenen op mondiale standaarden is onderdeel van de bredere strategie van het land, zoals we hierboven hebben gezien. Het land heeft inmiddels

816 Kaiser en Schot 2014.
817 Rühlig 2020: 11.
818 Cihon 2019: 10.

hooggeplaatste functionarissen geleverd voor de ISO, de IEC en de ITU. Op het niveau van de commissies en de werkgroepen binnen die organisaties groeit het aandeel Chinese afgevaardigden. Het absolute aantal is nog steeds lager dan dat van veel andere landen, maar het Chinese aandeel groeit wel.[819] Ook het *Belt and Road Initiative,* China's grote internationale project, heeft een expliciete component van standaardisering.[820] Ten slotte is er een geplande hervorming van het Chinese systeem naar aanleiding van de verkenning *China Standards 2035*.[821] Specifiek rondom AI groeit de Chinese invloed in de ITU, waar het land standaarden voor gezichtsherkenning en surveillance beïnvloedt. Chinese bedrijven als ZTE, Dahua en China Telecom leveren er voorstellen voor internationale standaarden. Door in dat forum standaarden te zetten, krijgt het land vooral invloed in opkomende markten in Afrika, Azië en Latijns-Amerika, die deze standaarden vaak overnemen. Het land versterkt zo de toegang tot voor hem belangrijke afzetmarkten voor deze technologieën. Op het gebied van surveillancestandaarden zijn Chinese voorstellen bijvoorbeeld gelijk aan het ontwerp van ZTE's *Smart Street 2.0* verkeerslicht.[822]

China is overigens niet de enige partij die bijdraagt aan de 'geopolitisering' van standaarden. Ook de VS doet daaraan mee, bijvoorbeeld op het gebied van telecommunicatiestandaarden in het handelsconflict met China.[823] Er wordt in die zin ook wel gesproken van 'connectiviteitsoorlogen'.[824] In het subcomité voor AI-standaarden van de ISO/IEC Joint Technical Committee for IT strijden de VS en China om het leiderschap.[825]

Als gevolg van deze ontwikkelingen staan de Europese rol en werkwijze – een relatief technocratisch proces gedreven door private partijen met expertise – bij het zetten van standaarden onder druk.[826] Deze ontwikkelingen vragen daarom om een antwoord van de Europese landen, en dus ook van Nederland, op de nieuwe politiek van standaardisering.

819 Rühlig 2020: 22.
820 In 2015 publiceerde de Chinese overheid het 'Action Plan for Harmonisation of Standards along the Belt and Road'. Als eerste stap daarbij werden 500 Chinese nationale en sectorale standaarden in andere talen vertaald (Rühlig 2020: 24).
821 Rühlig 2020: 18.
822 Gross et al., 1 december 2019
823 Voor een analyse van het belang van standaarden in relatie tot infrastructuur, zie: Hillman 2019.
824 Leonard 2016.
825 Smuha 2019: 21.
826 Cihon 2019.

Kernpunten – Van competitie naar samenwerking

- Nu geen land in staat zal zijn om AI volledig binnen de eigen grenzen af te bakenen en verder te ontwikkelen, is Nederland gebaat bij internationale samenwerking, in het bijzonder binnen de EU.
- Op vijf domeinen kan een land als Nederland de concurrentiekracht versterken door met geïntegreerd beleid van 'AI-diplomatie' in te zetten op internationale samenwerking.
- Op het gebied van fundamenteel onderzoek dragen projecten als ELLIS en het in Nederland gevestigde CLAIRE bij aan de Nederlandse positie.
- Op het gebied van regulering kan de EU haar marktmacht inzetten en een *first-mover advantage* creëren via regelgeving, waar Nederland weer indirect van kan profiteren.
- Nederland kan ook een actievere rol vervullen binnen internationale organisaties op het gebied van ethische richtlijnen en principes.
- Ten slotte vraagt de geopolitisering van standaardisatieprocessen om een antwoord van Nederland en andere Europese landen.

8.2 AI en nationale veiligheid

Het positioneren van een land op het gebied van AI heeft niet alleen betrekking op het verdienvermogen van dat land. De internationale dimensie gaat bij uitstek ook over vraagstukken betreffende veiligheid en nationale soevereiniteit. Verdienvermogen en veiligheid kunnen met elkaar verbonden zijn, zoals we zagen, maar met deze paragraaf richten wij ons primair op de vraagstukken rondom veiligheid. Onze focus ligt daarbij op nationale veiligheid en niet op de veiligheid van AI-toepassingen voor bijvoorbeeld burgers. Die thematiek is nader uitgewerkt in het eerdere WRR-rapport *Veiligheid in een wereld van verbindingen*.[827]

De invloed van AI op internationale machtsverhoudingen en daarmee op de nationale veiligheid wordt breed erkend. Beroemd is het citaat van de Russische president Vladimir Poetin uit een speech voor studenten en scholieren in 2017: "Het land dat leidt in AI is het land dat de wereld zal regeren." Ook in de VS wordt AI vanuit die optiek bekeken. Historisch heeft het leger daar een sterke band met Silicon Valley via organisaties als DARPA en investeringen van het Department of Defense.[828] In 2015 werd de *Defense Innovation Unit* (DIU)

827 WRR 2017.

828 DARPA is met ARPANET verantwoordelijk geweest voor de voorloper van het internet en heeft bijgedragen aan de ontwikkeling van allerlei geavanceerde wapens. Voor een historisch overzicht van de ontwikkeling van DARPA en het succes en falen van de organisatie, zie: Weinberger 2019.

opgericht om technologieën uit Silicon Valley beter in te kunnen zetten voor het leger. Specifiek ten aanzien van AI schreef toenmalig minister van Defensie Jim Mattis in 2018 een memorandum voor een integrale nationale strategie voor AI. Later dat jaar tekende president Trump de *National Defense Authorisation Act* (NDAA) waarmee ook de *National Security Commission for AI* werd opgezet. Het Pentagon zette ook het *Joint Artificial Intelligence Center* (JAIC) op.[829] In maart 2021 verscheen het dikke eindrapport van de *National Security Commission on Artificial Intelligence* (NSCAI), onder voorzitterschap van de hier al vaak genoemde oud-CEO van Google, Eric Schmidt.[830]

Op economisch terrein hebben wij kanttekeningen geplaatst bij het idee dat er een mondiale race om AI gaande is. Wat betreft de nationale veiligheid lijkt dat beeld adequater. Militaire capaciteiten bieden namelijk wel degelijk *zero sum*-voordelen aan landen. Dit wordt internationaal ook opgemerkt: terwijl er tot 2016 nog minder dan 300 onlinehits waren voor '*AI arms race*', groeide dat aantal in 2018 uit tot 50.000 hits en schrijven kranten als *The Guardian* en de *Wall Street Journal* uitgebreid over het fenomeen.[831]

Wanneer het gaat over de invloed van AI op veiligheid, gaat vaak als eerste over autonome wapens. Dat is een in het oog springend onderwerp waar veel om te doen is. We staan daarom eerst bij dit fenomeen stil, waarna we de focus zullen verbreden naar andere minder bekende manieren waarop AI de veiligheid beïnvloedt.

Autonome wapens

Autonome wapens spreken enorm tot de verbeelding. Ze zijn het onderwerp van veel dystopische literatuur. Het thema *'rise of the machines'* neemt vaak de vorm aan van robots die besluiten de aanval in te zetten op de mensheid. In hoofdstuk 4 over demystificatie hebben we laten zien dat de angst dat robots zelfbewust worden en besluiten dat mensen hun vijanden zijn, niet erg reëel is. Tegelijkertijd zagen we ook dat robots niet zelfbewust hoeven te zijn om een dreiging voor ons te vormen. Wat gebeurt er momenteel op het gebied van autonome wapens en wat voor implicaties heeft dat voor de internationale orde, en daarmee voor de veiligheid van Nederland?

Het is niet gemakkelijk om te definiëren wat een autonoom wapen is. Voordat we ons op dat vraagstuk kunnen richten, is het goed om een aantal bestaande technologieën op dit gebied te behandelen. De diversiteit aan systemen is

829 Leung 2019:266.
830 National Security Commission of Artificial Intelligence 2021.
831 Leung 2019: 263.

namelijk erg groot. De Israëlische Harpy-drone kan autonoom door de lucht vliegen op zoek naar radarsystemen van de vijand en die zonder toestemming te vragen zelf neerschieten. China heeft een eigen variant hiervan gemaakt met behulp van *reverse engineering*. In de gedemilitariseerde zone op de grens met Noord-Korea heeft Zuid-Korea een robotwapen dat autonoom kan schieten op bewegende objecten. Rusland bouwt gewapende grondrobots voor conflicten op de Europese vlakten. Ten minste dertig landen hebben defensiesystemen die, wanneer zij ingeschakeld zijn, autonoom inkomende gevaren als raketten kunnen neerhalen. De VS heeft daarvoor het AEGIS-systeem voor schepen en de Patriot voor raketten op het land. Andere voorbeelden zijn het Duitse MANTIS, het Israëlische Trophy en het Russische Arena-systeem.[832] Specifiek op het gebied van gewapende drones waren er in 2017 zestien landen die deze in bezit hadden. Van de internationale verkoop was 90 procent afkomstig van China.

Tekstbox 8.3 – Drones en oorlogsvoering

Drones werden in oorlogsvoering voor het eerst al gebruikt in de Vietnamoorlog, maar na 11 september 2001 nam het gebruik ervan een vlucht en zijn ze door het Amerikaanse leger veelvuldig ingezet in Afghanistan. In 2018 bleken ook Syrische rebellen in staat om een grote aanval met dertien drones op een Russische vliegbasis uit te voeren. Later dat jaar bracht Rusland het zwaarbewapende systeem Uran-9 naar dat slagveld. In augustus van dat jaar werd in Venezuela een poging gedaan om president Maduro te doden met een drone.[833]

Er gebeurt dus veel op het gebied van de ontwikkeling van autonome wapens. Het fenomeen maakt veel los en er zijn verschillende internationale acties geweest om dergelijke wapens te laten verbieden (zie hoofdstuk 6). Tegelijkertijd zijn er de nodige obstakels om tot een dergelijk verbod te komen.[834] Verschillende van die obstakels zijn illustratief voor bredere problemen met het mondiaal reguleren van AI en dus voor de opgave van positionering. Om die reden behandelen we die hier uitgebreider. Het gaat om het probleem van definities, het *dual use*-karakter van AI en het motiveren van landen om aan een verbod mee te doen.

In hoofdstuk 1 zagen we hoe moeilijk het is om AI te definiëren. Hoewel autonome wapens een vrij concreet toepassingsgebied van AI zijn, zijn ook die zeer

832 Scharre 2018: 45-46.
833 Scharre 2018: 364.
834 Zie in meer detail: Buruma 2020: 69-112.

moeilijk te definiëren en dat levert moeilijkheden op voor de regelgeving. Paul Scharre laat zien dat er weinig eenduidigheid is over autonome wapens, omdat autonomie langs drie verschillende assen gedefinieerd kan worden.[835] De eerste as betreft de aard van de taken. Sommige taken worden door mensen gedaan en andere door een machine. Een volledig zelfrijdende auto zal alles zelf kunnen doen, maar huidige auto's vervullen vaak al allerlei deeltaken zelf, zoals cruise control, inparkeren of remmen. Bij het laatste kan de auto de controle overnemen van de menselijke bestuurder in het geval van een potentieel ongeluk. Maar wanneer is een wapen nu autonoom? Is het dat al als het zelf door de lucht kan vliegen en andere objecten autonoom kan ontwijken, maar nog niet schiet?

Specifiek ten aanzien van de taak van het schieten kunnen we de vraag stellen of een systeem autonoom is als de mens op een knop drukt om in aanvalsmodus te gaan en de robot vervolgens de kogels zelf afvuurt. Er zijn ook technologieën waarbij een mens een object markeert om te vernietigen waarna de robot doorgaat totdat dat object vernietigd is. Is dit een autonoom wapen? Een manier om autonoom handelen te beperken is bijvoorbeeld om het alleen toe te staan voor taken van defensieve aard. Er zijn bijvoorbeeld systemen die automatisch een raketsalvo uit de lucht kunnen halen, een functionaliteit die veel mensen nuttig zullen vinden. Maar is het ook defensief als dat systeem terugvuurt naar de bron van die raketten? Wanneer dit systeem naar het grondgebied van de tegenstander wordt verplaatst, wordt het nog moeilijker te onderscheiden wat offensief en wat defensief gebruik van een autonoom wapen is.

Een tweede as waarlangs autonomie gedefinieerd kan worden, is de rol van de mens. Hiervoor wordt in het militaire domein hetzelfde kader gebruikt als wij in hoofdstuk 5 hebben besproken. Semi-autonome systemen zijn het equivalent van *human-in-the-loop*, waarbij een machine nadat deze een taak heeft vervuld, wacht op menselijke input om verder te gaan. Bij gesuperviseerde autonome systemen blijft een mens *on the loop* en kan die ingrijpen of het systeem stoppen. Bij volledig autonome systemen ten slotte is de mens *out of the loop* en heeft hij geen rol bij de beslissingen die het systeem neemt.

De derde as waarlangs autonomie onderscheiden kan worden, is het niveau van intelligentie. Een klassieke mijn handelt in principe autonoom door te ontploffen als iemand erop gaat staan. Toch zullen weinig mensen die mijn als een autonoom wapen beschouwen. Dat komt omdat het niveau van vereiste intelligentie laag is. Er hoeven geen complexe afwegingen gemaakt te worden voordat de mijn ontploft. Een betere duiding is dan ook dat de mijn automatisch is. Op een niveau hoger wordt wel een aantal variabelen meegewogen voordat tot actie

835 Scharre 2018: 27-32.

wordt overgegaan. Dat kunnen we geautomatiseerd noemen. Pas wanneer de activiteit veel complexer wordt, en een systeem eigenstandig afweegt op welke manier het een doel bereikt, zouden we kunnen spreken van autonomie.

Dat autonome wapens langs drie assen gedefinieerd kunnen worden, maakt het niet gemakkelijker om hierover internationaal tot overeenstemming te komen en om te bepalen wat een land als Nederland kan ambiëren op dit gebied. In tegenstelling tot andere soorten wapens spelen hier bijvoorbeeld vragen over het toekennen van autonomie aan defensieve systemen, aan systemen waar mensen algemene opdrachten aan geven, aan systemen waar een mens alleen controleert of aan systemen waar het niveau van intelligentie niet heel hoog is.

Naast het probleem van definities is er nog een probleem bij de internationale afstemming over autonome wapens dat eveneens een breder AI-vraagstuk illustreert. Dat is het *dual use*-karakter ervan. Daarmee wordt bedoeld dat er toepassingen zijn die tegelijkertijd vreedzaam civiel gebruikt kunnen worden, maar ook als wapen kunnen worden ingezet in een conflict.

Een voorbeeld hiervan is de *Fast Leightweight Autonomy* (FLA) van DARPA. Filmpjes van deze technologie zijn wereldwijd viraal gegaan. Begeleid met James Bondmuziek is een zwerm drones te zien die door ramen huizen binnenvliegen en allerlei complexe formaties in de lucht maken. De filmpjes hebben voor veel ophef gezorgd en FLA is dan ook een technologie waarop de beweging tegen autonome wapens zich richt. FLA is vooralsnog niet uitgerust met wapens. Het is echter duidelijk dat zeer goed vliegende drones die in formaties om objecten kunnen bewegen, voor het leger interessant zijn. Zolang er geen wapens in het spel zijn, is de achterliggende technologie van FLA echter vergelijkbaar met die van de zelfrijdende auto. Het gaat om lokalisatie, mapping, objectdetectie en dynamische navigatie op hoge snelheid.[836] Zelfrijdende auto's worden vanwege die technologische toepassingen echter niet als dreiging voor de veiligheid gezien.

Of neem het vermogen van een drone om zich op een object te richten en het vervolgens door de ruimte te volgen. Dat kan voor het leger heel nuttig zijn, bijvoorbeeld om drones een bewegende truck te laten achtervolgen. Diezelfde technologie heeft ook hele onschuldige toepassingen. Verschillende commerciële drones kunnen dit namelijk al en worden bijvoorbeeld gebruikt om een bruidsstoet te volgen en te filmen.

836 Scharre 2018: 70.

Wat autonome wapens ook gevaarlijk kan maken, is het vermogen om menselijke gezichten te detecteren of specifieke mensen te identificeren. Gecombineerd met een geweer is dit iets heel gevaarlijks. De software hiervoor wordt echter al in allerlei andere toepassingen gebruikt. In grote open source software-databases zoals TensorFlow is die software zelfs vrij beschikbaar en wordt hulp aangeboden om algoritmes op die taken te trainen.

De onderliggende technologie van autonome wapens, van navigatie en het volgen van objecten tot gezichtsherkenning, wordt dus ook in allerlei commerciële vreedzame toepassingen gebruikt. Dat maakt iets als een algemeen verbod moeilijk. Een drone die iemands gezicht herkent om die vervolgens diens pakketje te leveren, bevat bijna dezelfde technologie als de drone die diezelfde persoon kan neerschieten. Daar komt nog bij: wanneer gezichtsherkenning niet wordt ingezet om mensen te doden, maar juist om mensen te vermijden, is gezichtsherkenning dan ook verwerpelijk? Zelfs als dit de kans op burgerslachtoffers kan verminderen? Kortom, de brede inzetbaarheid van AI maakt het veel moeilijker om het militaire gebruik ervan aan banden te leggen dan bij bijvoorbeeld chemische wapens of langeafstandsraketten.[837]

Tekstbox 8.4 – Visies over het effect van autonome wapens

Verschillende AI-wetenschappers verwachten dat de inzet van autonome wapens oorlog juist minder destructief maakt. Nick Bostrom benadrukt het belang om mensen zoveel mogelijk uit het slagveld te verwijderen en de mogelijkheid van minder dodelijke slachtoffers door de precisie van autonome wapens.[838] Rodney Brooks geeft aan dat in tegenstelling tot een mens een robot het zich kan permitteren om pas te schieten als erop geschoten wordt.[839] Volgens Yann LeCun veranderen de grotere precisie en het mogelijk minder dodelijke karakter van deze wapens het leger meer in zoiets als een politiedienst.[840] Dat klinkt als een positieve ontwikkeling, maar volgens Frank Pasquale brengt ze ook nieuwe gevaren. Als strijd meer het karakter van een internationale politieactie krijgt, waarbij minder slachtoffers vallen, zullen politici ook minder druk voelen om soldaten te sparen en zullen oorlogen sluimerend kunnen worden en minder gemakkelijk te beëindigen zijn.[841]

837 Dat betekent dat om gevaren te adresseren meer gekeken zou moeten worden naar praktijken bij andere *dual use*-technologieën (Brundage et al. 2018).
838 Uit een interview met Nick Bostrom (Ford 2018: 107).
839 Uit een interview met Rodney Brooks (Ford 2018: 440).
840 Uit een interview met Yann LeCun (Ford 2018: 137).
841 Pasquale 2020: 152.

Het derde probleem bij het komen tot eenduidige internationale afspraken over autonome wapens betreft de motivatie van sterke landen. Een zekere internationale consensus bestaat inmiddels rondom het idee van 'betekenisvolle menselijke controle'.[842] Het Nederlandse kabinet reageerde met een uitgebreid standpunt[843] en bevestigde naar aanleiding van vragen van het Tweede Kamerlid Van der Staaij eind maart 2018 nogmaals dat "betekenisvolle menselijke controle altijd noodzakelijk is bij de inzet van autonome wapensystemen."[844] De ministers Blok en Bijleveld merkten daarbij op dat "Nederland als een van de weinige landen met een uitgebreid kabinetsstandpunt over dit onderwerp de samenvatting hiervan ook als *non-paper*" ingediend heeft bij het debat over autonome wapensystemen dat plaatsvindt binnen de *Convention on Certain Conventional Weapons* (CCW) van de VN.

Het zal echter toch lastig blijven om met name landen met geavanceerde legers te verhinderen om wapens met meer en meer autonomie te ontwikkelen. Er worden, zoals gezegd, sterke internationale campagnes tegen dergelijke wapens gevoerd. Opvallend daaraan is dat die goeddeels gedreven worden door ngo's en minder door staten. Sterker nog, geen enkel sterk land schaart zich vooralsnog achter een compleet verbod op deze wapens. Landen die zich daar wel voor hebben uitgesproken, zijn landen als Ecuador, Ghana, Irak en Pakistan. Er zitten ook landen bij die geen legers hebben, zoals Costa Rica en het Vaticaan. Het zijn met andere woorden geen militaire grootmachten en ook niet de landen die vooroplopen bij de bescherming van mensenrechten. Het betreft vooral landen die vrezen wat sterkere landen ermee kunnen doen en die met een verbod hopen hun handen te binden.[845] Hoewel sommige Europese landen zoals Oostenrijk ook die kant op neigen, is het zonder de steun van landen met sterke legers erg lastig om een verbod daadwerkelijk in te voeren.

Scharre merkt bovendien op dat een internationaal verdrag niet doorslaggevend is voor het tegengaan van het gebruik van bepaalde wapens. Er zijn gevallen van wapens die, ondanks een verbod, toch gebruikt zijn en er zijn ook gevallen van wapens die niet gebruikt zijn zonder dat er een verdrag over was. Belangrijk hierbij is vooral de vraag of landen wederkerigheid verwachten. De angst dat een ander land het wapen ook kan gebruiken, werkt afschrikkend. Landen die dat niet vrezen van een ander land, zijn minder geremd om nieuwe technologieën te gebruiken.[846] Wat landen verder afschrikt, is transparantie over het gebruik van een bepaald wapen, waardoor het ter verantwoording geroepen kan

842 AIV en Commissie van Advies Inzake Volkenrechtelijke Vraagstukken 2015.
843 Kamerstukken II 2015/16, 34300-X, nr. 88.
844 Aanhangsel Handelingen II 2017/18, nr. 1645.
845 Scharre 2018: 350.
846 Scharre 2018: 340.

worden. Dat is in het geval van autonome wapens lastig. De autonomie schuilt namelijk niet in de hardware of bepaalde duidelijke fysieke kenmerken, maar in de software. Daarmee is het een stuk lastiger om transparantie te geven over wat er in een conflict gebeurt.

De drie behandelde vraagstukken over het definiëren van autonome wapens, de verwevenheid met civiele technologieën en de motivatie van sterke landen, tonen de complexiteit van internationale coördinatie bij deze technologie. Dat wil niet zeggen dat coördinatie onmogelijk is. Deze casus is illustratief voor obstakels die spelen rondom de internationale afstemming op AI-gerelateerde vraagstukken. Het eerder genoemde rapport van de AIV laat zien welke rol Nederland hierbij kan spelen.

De discussie over de invloed van AI op veiligheid gaat, zoals gezegd, vaak over autonome wapens. Omdat deze zo tot de verbeelding spreken, krijgen andere toepassingen onterecht minder aandacht. AI kan oorlogsvoering namelijk ook op andere manieren beïnvloeden (zie hieronder).

Kernpunten – Autonome wapens

- Ten aanzien van AI en veiligheid gaat veel aandacht uit naar autonome wapens, die tot de verbeelding spreken en veel verzet oproepen. Verschillende obstakels maken het moeilijk om op dit dossier tot heldere internationale afspraken te komen.
- Autonome wapens kunnen langs drie assen gedefinieerd worden, wat een algemene definitie in de weg staat en daardoor regelgeving bemoeilijkt.
- Het *dual use*-karakter van veel van de technische benodigdheden maakt controle of een verbod erg lastig.
- Landen met sterke legers verzetten zich tegen beperkingen op dit terrein.
- Nederland moet met die obstakels rekening houden in haar positionering om de internationale veiligheid te versterken en de nationale veiligheid te borgen.

Andere militaire toepassingen

In een studie voor de NAVO naar de invloed van AI op oorlogsvoering onderscheidt Matej Tonin naast autonome robotsystemen ook het veld van ondersteuning in informatievoorziening en besluitvorming.[847] AI kan de snelheid van analyse en besluitvorming verhogen. Dat kan door de reactietijd van defensieve systemen te verhogen, door nuttige informatie aan te leveren aan besluitvormers – wat een voordeel op kan leveren ten opzichte van rivalen –, door het snel detecteren van cyberaanvallen en door het identificeren van pogingen om desinformatie te verspreiden. Niet alleen de snelheid, maar ook de kwaliteit van besluitvorming kan door AI verbeterd worden. Tonin citeert hierbij een Britse officier die een situatie constateerde van "zwemmend in sensoren, verdrinkend in data en hongerend naar inzicht". Hier kan AI helpen door bijvoorbeeld de analyse van surveillancedata, het uitlichten van abnormale patronen of het opvangen van zwakke signalen van mogelijke dreigingen. De op de lange termijn gerichte Defensievisie die minister Bijleveld in oktober 2020 presenteerde, spreekt in dit verband van 'informatiegestuurd optreden' als basis van de toekomstige defensieorganisatie.[848]

Het is hier relevant om op te merken dat de laatste jaren verschillende belangrijke onthullingen over militaire organisaties uit heel basale informatiebronnen kwamen. De data van de hardloopapp Strava onthulde de locatie van een geheime Amerikaanse militaire basis in Afrika. Een geheim Chinees schip werd onthuld op de achtergrond van een foto van een toerist.[849] Met satellietdata ontdekten wetenschappers in 2021 silo's voor kernwapens in China. AI zou mogelijk een flinke bijdrage kunnen leveren aan het analyseren van dergelijke basale informatiebronnen voor militair relevante data[850].

Het gebruik van AI voor informatievoorziening brengt ook bredere vraagstukken rondom AI met zich mee in het domein van defensie. Een eerste vraagstuk is het fenomeen van '*novelty detection*'.[851] Het gaat daarbij om het vermogen van AI-systemen om te detecteren dat de input die zij krijgen, buiten het bereik valt van waar ze op getraind zijn. Een algoritme getraind om honden te onderscheiden, moet op een afbeelding een hond kunnen onderscheiden die het

847 Tonin 2019.
848 Ministerie van Defensie 2020.
849 Singer en Brooking 2018: 58.
850 Kate Crawford geeft een treffend voorbeeld van de afleiding van allerlei persoonlijke informatie. In 2013 werd een dataset van 173 miljoen geanonimiseerde taxiritten uit New York vrijgegeven. Onderzoekers hebben de data toen snel gedeanonimiseerd, de jaarinkomsten van chauffeurs berekend, een aantal beroemde mensen geïdentificeerd die stripclubs hadden bezocht en afgeleid welke chauffeurs moslim waren door te kijken naar pauzes tijdens gebedstijden (Crawford 2021: 111).
851 Lin 2019: 145.

nooit eerder heeft gezien. Als het een afbeelding van een dolfijn krijgt, moet het echter kunnen onderscheiden dat het zoveel van honden verschilt dat het iets totaal nieuws is, en niet te categoriseren is als het type hond waar de dolfijn het meest op lijkt. Het algoritme moet dus een onderscheid maken tussen 'verschillend in detail, maar heel vergelijkbaar' en 'zeer onvergelijkbaar'. In het militaire domein is dit belangrijk om ervoor te zorgen dat enerzijds elk type aanvalsraket wel als zodanig wordt geïdentificeerd, maar dat een vliegtuig bijvoorbeeld nooit tot die categorie wordt gerekend.

Een tweede vraagstuk betreft de manipulatie van data om een tegenstander te misleiden. Zo kan gezocht worden naar '*edge cases*', gevallen waarin een zwakte van een algoritme tot een extreem andere uitkomst leidt. In laboratoria hebben onderzoekers afbeeldingen van bussen subtiel gemanipuleerd om als struisvogels te worden gemarkeerd en schildpadden als geweren. Bij een ander onderzoek leidde het veranderen van een paar pixels ertoe dat een neuraal netwerk een foto van een olifant als een auto identificeerde.[852] Wij zagen in hoofdstuk 5 al hoe dit problemen kan opleveren voor zelfrijdende auto's. Dergelijke manipulatie is in het militaire domein nog gevaarlijker. Systemen kunnen misleid worden om aanvallen niet te herkennen. Omgekeerd kunnen ook zaken die niet dreigend zijn wel als zodanig gepresenteerd worden om een aanval uit te lokken. We kunnen ons hierbij hele complexe situaties voorstellen waarbij bijvoorbeeld strijders op heel subtiele wijze een ziekenhuis markeren als een militaire installatie waardoor het wordt aangevallen. Als die misleiding heel klein en subtiel is, valt dit moeilijk te achterhalen en is het vervolgens nauwelijks mogelijk om te bewijzen dat het niet de intentie was om het ziekenhuis te bombarderen.[853]

Als systeemtechnologie is de impact van AI op het militaire domein niet goed te voorspellen. Ze kan op zeer verschillende wijzen worden toegepast. Dat is ook een van de redenen waarom legers zo in de technologie geïnteresseerd zijn: om ervoor te zorgen dat zij niet een groot strategisch voordeel missen. Het is daarom belangrijk om in dit gebied verder te kijken dan alleen autonome wapens. Bovendien komen bedreigingen van de veiligheid en de nationale soevereiniteit niet alleen voort uit innovatie in het leger. Ook niet-militaire ontwikkelingen moeten goed in de gaten worden gehouden en kunnen van invloed zijn op veiligheidskwesties, zoals het voorbeeld van Strava laat zien.

852 Libicki 2019: 139-140.
853 Lin 2019: 149.

Kernpunten – Andere militaire toepassingen
- AI kan bijdragen aan de snelheid en kwaliteit van analyse en besluitvorming op het militaire toneel.
- Het gebruik van AI voor informatievoorziening brengt nieuwe technische vraagstukken met zich mee zoals *novelty detection* en *edge cases*.
- AI kan ook allerlei militaire processen ondersteunen.
- Als systeemtechnologie is de impact van AI op oorlogsvoering uiteindelijk niet te voorspellen, waardoor het van belang is om scherp te zijn op de waarde van AI op militair gebied.

Veiligheid buiten het strijdtoneel

Binnenlandse en buitenlandse veiligheid raken steeds meer met elkaar verweven.[854] De laatste jaren is er steeds meer aandacht voor cyberaanvallen op de infrastructuur van samenlevingen. Denk aan de opkomst van ransomware en de cyberaanval op Oekraïne. Dat sorteerde wereldwijd grote neveneffecten, onder andere op containerbedrijf Maersk, met problemen voor de Rotterdamse haven als gevolg. De WRR heeft in 2019 een rapport gepubliceerd over dit risico op digitale ontwrichting.[855] De groei van sensoren in fysieke objecten, het IoT, brengt nieuwe kwetsbaarheden mee voor aanvallen met AI.[856]

Naast de infrastructuur van de digitale wereld is ook steeds vaker de stroom van informatie in de digitale wereld de inzet van conflict en zorgen over veiligheid. We noemden al hoe apps en foto's van toeristen veiligheidsimplicaties kunnen hebben. Allerlei andere vormen van reguliere informatie kunnen van belang zijn in de competitie tussen landen. Bijvoorbeeld de informatie die mensen delen op sociale media. Het is bekend dat Russische geheime diensten de tweets van president Trump analyseerden om een psychologisch profiel te maken.[857] De informatie op sociale media van leiders, ambtenaren of gewone burgers kan waardevolle informatie bevatten voor buitenlandse rivalen.

AI kan niet alleen snel veel van dat soort informatie analyseren, het kan er ook nieuwe patronen uit destilleren. Onderzoek door platformen als Facebook heeft psychologische inzichten over mensen aan het licht gebracht. Zo wordt beweerd dat het consistent gebruik van zwart-witfilters op Instagram en het plaatsen van foto's met alleen een gezicht indicatoren zouden zijn voor klinische depressie.

854 WRR 2017.
855 WRR 2019.
856 Brundage et al. 2018.
857 Singer en Brooking 2018: 61.

Meer en meer onderzoek haalt medische data uit videobeelden van mensen.[858] Het is dan ook niet onwaarschijnlijk dat verschillende landen zoveel mogelijk informatie van de sociale media uit andere landen, waaronder Nederland, verzamelen en daar zoveel mogelijk algoritmes op loslaten.

Naast het verzamelen van waardevolle informatie is ook het beïnvloeden en sturen van informatie een slagveld geworden in wat Singer en Brookings de *LikeWar* op sociale media hebben genoemd. Naast militaire tactieken gebruikte de terreurorganisatie IS ook veelvuldig sociale media. Sterker nog, hun strijd werd daar evengoed gestreden als op het veld in Irak en Syrië. Via Twitter werd verslag gedaan van conflicten, werden nieuwe leden gerekruteerd met professionele actievideo's en werd verwarring gezaaid in de steden die zij aanvielen. Zoals de Nazi's de radio gebruikten voor snelle communicatie en om in Frankrijk verwarring te zaaien, wat bijdroeg aan het passeren van de Maginot-Lijn, zette IS in een eenentwintigste-eeuwse Blitzkrieg sociale media in.[859] Veel van dit gebruik van sociale media was gewoon mensenwerk, maar de grote invloed daarvan voert terug op de manier waarop algoritmes berichten viraal kunnen laten gaan. Militaire conflicten hebben tegenwoordig dus ook een digitaal 'slagveld' waar op de sociale media gestreden wordt om het juiste frame en de publieke opinie in het eigen land en in andere landen beïnvloed wordt.

AI en specifiek methoden van machine learning worden op allerlei manieren ingezet in deze 'informatieoorlog' (zie ook tekstbox 8.5).[860] Allereerst via *micro-targeting*, dat we reeds in hoofdstuk 2 bespraken. Het gaat daarbij om het maken van profielen van mensen om erachter te komen welke boodschappen het beste bij hen resoneren. Daarnaast wordt *sentiment analysis* en *natural language processing* (NLP) gebruikt om populaties beter te begrijpen en ze vervolgens bijvoorbeeld gerichte berichten te sturen. Een toepassing van NLP zijn chatbots die steeds beter mensen kunnen imiteren en zo gebruikt kunnen worden om hen te beïnvloeden.

Tekstbox 8.5 – Moderne propaganda

Waar propaganda vroeger verspreid werd door radioprogramma's in een ander land uit te zenden om zo invloed uit te oefenen op de meningen en gevoelens van mensen, gebeurt dat nu via sociale media. Nu kan zelfs direct met mensen contact worden opgenomen,

858 Het Chinese bedrijf Tencent werkt samen met een Brits zorgbedrijf om Parkinson te herkennen uit videobeelden van mensen (Ram, 7 mei 2019).
859 Singer en Brooking 2018: 7.
860 Kerr 2019: 71.

door ze vriendschapsverzoeken te sturen en ze vervolgens met informatie te overspoelen. Door middel van 'likes' kunnen burgers van ver weg participeren in een strijd en de propaganda eromheen. Onlinegeldinzamelingsacties maken die bijdrage nog concreter.

Democratische instituties kunnen door informatieoorlogen op het spel komen te staan. Bij de Amerikaanse presidentsverkiezingen van 2016 was er sprake van zowel binnenlandse als buitenlandse actoren, zoals de Russische staat, die poogden middels sociale media de mening van de bevolking te beïnvloeden. Jeff Giesea werkte voor de onlinecampagne van Trump. De gebruikte tactieken beschrijft hij in een paper voor een tijdschrift van de NAVO als *'memetic warfare'* – naar de memes die online viraal gaan.[861] Hij trekt daarin parallellen tussen de tactieken van de Trump-campagne en die van IS en Russische propagandisten.

Een andere oprukkende toepassing waar we iets uitgebreider bij stil moeten staan, zijn *'deepfakes'*. Het gaat hierbij om de productie van nepbeelden of geluidsfragmenten die moeilijk van echt te onderscheiden zijn. In 2017 presenteerde het bedrijf Lyrebird een choquerend echt lijkende opname van een gesprek tussen Barack Obama, Hillary Clinton en Donald Trump. Op het gebied van beeld hebben wetenschappers daarnaast een tweedimensionale foto van iemands gezicht omgezet in een driedimensionaal gezicht dat vervolgens tot bewegen gebracht kon worden. Met gewone camera's hebben weer andere wetenschappers eerst de gezichtsuitdrukkingen van een bepaald persoon tijdens het praten onderzocht en die vervolgens met die van een ander gecombineerd (*deformation transfer*). Daardoor kon een persoon op de eigen manier iets vertellen wat dan simultaan werd vertaald naar de uitdrukkingen van een andere persoon. Op die manier is ook beeld- en geluidsmateriaal van Barack Obama verzameld, waarvan er heel veel publiekelijk beschikbaar is, waardoor nu beeld gecreëerd kan worden waarin hij alles zou kunnen zeggen wat iemand hem in de mond wil leggen.[862] Lyrebird zegt nu nog maar een paar minuten trainingsdata nodig te hebben om realistische *deepfake*-audiofragmenten te maken.[863]

Nepbeelden betreffen ook het creëren van totaal nieuwe beelden. Hiervoor worden de in deel 1 besproken *generative adversarial networks* (GAN's) gebruikt. Voor Hollywoodfilms of computerspellen worden die gebruikt om nieuwe

861 Giesea 2016.
862 Singer en Brooking 2018: 254-255.
863 Schick 2020: 148.

objecten en omgevingen te genereren. Ze zijn ook gebruikt om nepgezichten te maken van niet-bestaande mensen. Die beelden zijn inmiddels zo goed dat veel mensen ze niet kunnen onderscheiden van echte foto's. Op die manier kunnen gebeurtenissen en handelingen van mensen op beeld gezet worden die nooit hebben plaatsgevonden, maar die niet van echt te onderscheiden zijn. Niet verrassend merkt demissionair minister van Defensie Bijleveld in de inleiding bij de Nederlandse Defensievisie 2035 op dat moderne conflicten die gepaard gaan met nepniews, beïnvloeding en polarisatie Defensie dwingen tot vernieuwen.[864]

Door de opkomst van technologieën om nepmateriaal te produceren voorziet technologieonderzoeker Aviv Ovadya iets wat hij de '*infocalypse*' noemt, een samenstelling van 'informatie' en Apocalyps.[865] Het fenomeen staat voor de combinatie *van deepfake*-beelden, chatbots die mensen heel goed kunnen imiteren (*laser phishing*) en allerlei andere vormen van manipulatie die het vrijwel onmogelijk zullen maken om nep en echt van elkaar te onderscheiden. Zelfs als het onderscheid achteraf wel gemaakt kan worden, kunnen nepbeelden die echt genoeg lijken, schade veroorzaken als mensen ze aanvankelijk geloven.

Hedendaagse ontwikkelingen als nepnieuws en gehackte twitteraccounts die valse berichten verspreiden, zijn een klein voorproefje van dit fenomeen van de infocalyps. De infocalyps kan maatschappelijke scheidslijnen verdiepen door groepen tegen elkaar op te zetten en het wantrouwen in instituties te vergroten, maar ook leiden tot apathie over berichtgeving als alles manipuleerbaar lijkt te zijn. De filosoof Daniel Dennett speculeert zelfs over het einde van het moderne tijdperk van fotografisch bewijs en een terugkeer naar een wereld waarin mensen meer leunen op herinnering en vertrouwen dan op onomstotelijk bewijs.[866] Deze technologieën kunnen ook gebruikt worden om critici te smoren. Van een Indiase journaliste die erg kritisch was op de regering is bijvoorbeeld een *deepfake* pornografische video gemaakt die mede door politici is verspreid.[867] Meer en meer landen zetten desinformatie ook in als wapen. Volgens onderzoekers uit Oxford voerden 28 landen in 2017 desinformatie-operaties uit, wat in 2020 al was gestegen naar 70 landen.[868] Als zodanig is dit fenomeen een bedreiging voor het functioneren van de democratie.[869] Dit sluit aan op een breder fenomeen van niet-militaire bedreigingen door AI en bespreken we daarom apart.

---

864 Ministerie van Defensie 2020.
865 Warzel, 11 februari 2018.
866 Dennett 2019: 46.
867 Schick 2020: 125.
868 Schick 2020: 85.
869 Zie hiervoor ook het interview met Dr. Hans-Jakob Schindler in: Semaan, 16 maart 2020.

Kernpunten – Veiligheid buiten het strijdtoneel
- Naast de infrastructuur is steeds vaker ook de informatie van de digitale wereld reden voor zorgen over veiligheid.
- AI kan kwetsbare informatie destilleren uit het steeds grotere aantal databronnen in het civiele domein.
- Statelijke en niet-statelijke actoren zetten ook in op het sturen en beïnvloeden van mensen in een informatieoorlog.
- Allerlei toepassingen van AI als *microtargeting* en *deepfakes* spelen een rol in de informatieoorlog, die een groeiende bedreiging is voor het democratisch bestel in landen als Nederland.

Digitale dictatuur

Ook relevant voor de internationale veiligheid is de vraag hoe een technologie als AI zich verhoudt tot verschillende politieke regimes. Lange tijd was het antwoord daarop simpel en geruststellend: moderne technologie decentraliseert en democratiseert. De complexiteit van moderne technologie werd eerder al gezien als een factor waardoor de centrale planning van de Sovjet-Unie moest falen. Gecentraliseerde regimes zijn niet in staat om innovatie te genereren en coördinatievraagstukken op te lossen.[870]

Zoals we in het vierde hoofdstuk zagen, was er rondom het internet vanaf het begin een sterke ideologie die ervan uiting dat het een kracht was die centrale macht tegenging en individuen bevrijdde van de grip van autoriteiten. De Arabische lente die in 2011 begon, werd ook toegeschreven aan de democratiserende werking van internetplatformen. Deze visie op digitale technologie is echter aan het kenteren. De auteur Evgeny Morozov stelt dat individuen en bedrijven simpelweg als eerste de weg naar digitale technologie hadden gevonden. Inmiddels hebben overheden die weg ook gevonden en blijken zij evengoed internettechnologieën te kunnen toepassen voor centralisatie, controle en surveillance.[871] Dat geldt natuurlijk voor de regimes van landen als Iran, Rusland en China, maar ook voor westerse veiligheidsdiensten als de Amerikaanse NSA.

Specifiek ten aanzien van AI zijn er redenen om te geloven dat de technologie zelfs centralisatie in de hand kan werken. AI biedt bijvoorbeeld de mogelijkheid om goedkoop en op grote schaal een bevolking te monitoren. Voorheen moest de Stasi een groot aantal mensen in dienst nemen om een deel van de bevolking te kunnen bespioneren. Daarmee was er een grens aan de omvang en activiteiten van de organisatie. Nu kunnen algoritmes het werk doen door

870 Zie het werk van Hayek (1944) en de analyse van Fukuyama (1992) over de val van de Sovjet-Unie.
871 Morozov 2011.

patroonanalyse op een veel grotere hoeveelheid data dan ooit tevoren. Dit hangt ook samen met het genoemde *dual use*-karakter van AI. Massasurveillance vraagt niet zoals voorheen om enorme aparte investeringen en personeel, maar kan in hoge mate voortbouwen op de capaciteiten die er al zijn voor civiele toepassingen. Dat maakt ook voor overheden de drempel lager om AI op deze manier in te zetten.[872] Simpelweg het bewustzijn van mogelijke surveillance kan bovendien bij burgers leiden tot *chilling effects* en zelfcensuur.

Voorheen was decentralisatie noodzakelijk om informatie te coördineren. De vrije markt geeft een immens aantal signalen rondom vraag en aanbod om tot een prijszetting te komen. Geen twintigste-eeuws planbureau kon dat beter. Als we nu kijken naar bijvoorbeeld navigatieapps, dan is het wel mogelijk voor een centrale partij om alle informatie te verzamelen en een optimale verkeersstroom op te zetten. De oprichter van het Chinese platform Alibaba, Jack Ma, stelt dan ook dat AI betere centrale planning mogelijk maakt. Met voldoende informatie kunnen planners de economie beter begrijpen, voorspellen en aansturen.[873]

Technologieën als AI bieden nieuwe mogelijkheden om menselijk gedrag te beïnvloeden, om ze te *nudgen,* ze op subtiele wijze een kant op te duwen. De vrees is dat niet alleen computers, maar ook mensen op die manier worden 'geprogrammeerd'.[874] De auteurs van een artikel in de *Scientific American* wijzen op het gevaar van een geautomatiseerde samenleving met totalitaire trekken – een digitale variant van George Orwell's *Big Brother.*[875]

Uiteraard is het zo dat technologieën als AI tegelijkertijd de democratie kunnen versterken. Naast de oudere ideeën over decentralisatie en participatie ontstaan ook steeds meer partijen die nieuwe instrumenten inzetten voor democratische doelen. GVA *Dictator Alert* is bijvoorbeeld een algoritme dat vliegtuiggegevens scant om aan te geven wanneer een dictator in Genève landt. In hoofdstuk 6 hebben we verschillende voorbeelden laten zien van partijen in het maatschappelijk middenveld die AI gebruiken om rechtvaardigheid en inclusie te bevorderen. Er zijn bovendien verschillende projecten waarbij AI wordt ingezet om de Sustainable Development Goals (SDG's) van de VN te realiseren door

872 Wright 2019a: 38
873 Wright 2019b: 28. Ook het Amerikaanse bedrijf Predata levert bijvoorbeeld analyses op basis van webmonitoring voor toekomstige brandhaarden en knelpunten.
874 Helbing et al. 2017.
875 Om het verschil te markeren met oude vormen van surveillance spreekt Shoshana Zuboff van de 'Big Other' (Zuboff 2019).

bijvoorbeeld de positie van kleine boeren in ontwikkelingslanden te verbeteren.[876] AI wordt ook ingezet om mensenrechtenschendingen te monitoren.[877]

Tegelijkertijd lijkt het momentum van niet-democratische effecten van AI te groeien, mede gedreven door de opkomst van een aantal autoritaire landen. In de politicologie wordt gesproken van drie historische golven van democratisering: de eerste van 1820 tot 1926, de tweede vlak na de Tweede Wereldoorlog en de derde vanaf 1975.[878] Elke golf is tot nu toe gevolgd door een terugslag of tegenbeweging. Nicholas Wright doet een interessante suggestie die hier relevant is. Elke terugslag kende namelijk een andere vorm van dictatuur. Na de eerste golf was dat het fascisme. De tweede werd gevolgd door 'bureaucratisch autoritarisme', een aanduiding voor bijvoorbeeld de dictaturen in Latijns-Amerika vanaf de jaren zestig. De derde golf zou volgens Wright gevolgd kunnen worden door 'digitaal autoritarisme'.[879]

Volgens de jaarlijkse studie van Freedom House is een dergelijke terugslag echter al langer bezig. De regimes van leiders als Duterte, Erdogan, Poetin, Orbán en Bolsonaro worden al een tijd aangeduid als 'onliberale democratieën', een concept waar nieuwe technologie geen centrale rol in speelt. Het is echter wel mogelijk dat digitale technologieën dit type autoritaire regimes meer en meer versterken.

Bovendien ontstaat nu niet alleen een vorm van dictatuur die op digitale technologie leunt, maar dit model van bestuur wordt in toenemende mate ook geëxporteerd. De technologieën en bestuursmodellen van deze digitale dictatuur worden geëxporteerd naar andere dictaturen en naar relatief zwakke democratieën in bijvoorbeeld Afrika of Latijns-Amerika en zelfs naar ontwikkelde democratieën in Europa. Steven Feldstein heeft een *AI Global Surveillance Index* ontwikkeld en laat zien dat ten minste 75 landen vormen van AI-surveillance gebruiken in smart city-platformen, gezichtsherkenningssystemen en *smart policing*. Chinese bedrijven spelen hierbij een centrale rol, maar ook bedrijven uit de VS (Cisco, IBM), Frankrijk (Thales, Teleste), Japan (NEC) en Duitsland (Bosch) dragen hieraan bij.[880]

Een rapport van Amnesty International uit 2020 over de export van Europese surveillancetechnologie bracht ook de activiteiten van een Nederlands bedrijf aan het licht. Het bedrijf Noldus Information Technology leverde het product

876 Hirsch Ballin 2021: 29-30.
877 Isha Salian 2019.
878 Huntington 1991.
879 Wright 2019b: 24.
880 Feldstein 2019.

FaceReader, dat gezichtsuitdrukkingen analyseert, voor het Chinese ministerie voor publieke veiligheid, een ministerie dat volgens het rapport biometrische data veelvuldig inzet voor massasurveillance.[881] De internationale proliferatie van dit soort technologieën is een probleem voor de internationale orde en de waarden die Nederland daarin hoog wil houden.

Om meer invulling te geven aan de aard en omvang van digitale dictatuur eindigen wij dit hoofdstuk met twee casestudies. Daarbij zullen wij het fenomeen nader onderzoeken in China en Rusland, de landen met de meest geavanceerde capaciteiten.

Kernpunten – Digitale dictatuur
- Technologieën als AI kunnen een decentraliserende en democratiserende werking hebben. Daarnaast slagen autoritaire regimes er ook in toenemende mate in om dergelijke technologieën voor hun doelen in te zetten. Zo wordt er nu gesproken van 'digitale dictatuur'.
- In toenemende mate worden de instrumenten daarvan geëxporteerd, wat wereldwijd democratieën onder druk zet.
- Niet alleen autoritaire regimes, maar ook westerse, en zelfs Nederlandse, bedrijven dragen daaraan bij.

Casus: Digitale dictatuur in China

We hebben in paragraaf 8.1 gezien dat China AI heeft omarmd en middels het AIDP mondiaal leiderschap nastreeft. Het land zet de technologie ook expliciet in voor de surveillance en controle van de eigen bevolking.

Hierbij moet opgemerkt worden dat hoewel AI hier een belangrijk onderdeel van is, zij samengaat met een veel bredere set technologieën en (niet-digitale) praktijken. Belangrijk is bijvoorbeeld het *grid management*-systeem, waarbij gridmanagers de verantwoordelijkheid hebben om informatie te verzamelen over een deel van een buurt. Het *Golden Shield Project* levert bredere digitale technologie voor het bestuur van de bevolking en coördinatie binnen de overheid. Dataverzameling gebeurt ook middels een groot netwerk van sensoren in de fysieke omgeving, dat aangeduid wordt met plannen voor het *Internet Plus*.[882]

Een andere component van de Chinese dataverzameling is het programma *SkyNet* (bijzonder genoeg dezelfde naam als de kwaadaardige machine die zich

881 Amnesty International 2020b: 29.
882 Hoffman 2019: 51-52.

tegen de mensheid keert in *The Terminator*), dat erop gericht is om een nationaal netwerk van CCTV-camera's te installeren. In 2010 hingen er in Beijing al 800.000 camera's en in 2015 claimde de politie dat 100 procent van de stad was gedekt. Dat jaar plande het overheidsorgaan NDRC om in 2020 alle publieke ruimtes en leidende industrieën met een surveillancesysteem getiteld *Sharp Eyes* te dekken.

Bij dergelijke immense databronnen is het nodig om AI te gebruiken voor analyse. Bedrijven als Hikvision, Sensetime, Yitu en Megvii ontwikkelen slimme camera's hiervoor. SenseTime heeft bijvoorbeeld als doel om 100.000 hogeresolutievideofeeds tegelijkertijd te kunnen monitoren en door die bronnen individuen *in real time* te kunnen identificeren en volgen. In 2018 kon de politie zo een voortvluchtige identificeren en aanhouden bij een concert met 60.000 bezoekers.[883] Gezichtsherkenning wordt in China breed gebruikt.[884]

Centraal bij het aansturen van de Chinese bevolking is ook het beroemde socialekredietsysteem. Dat is vooralsnog niet een enkel systeem, maar bestaat uit een set van verschillende regionale en nationale projecten. Op nationaal niveau bestaat bijvoorbeeld het Xinyi+-Project, waarbij bedrijven als Ant Financial (financiële zaken), Didi Chuxing (Chinese Uber) en Ctrip (reisboekingen) samenwerken op het gebied van transport en verhuur om bepaalde mensen te weren en voor anderen juist een groter gemak te creëren naar gelang hun score. Een voorbeeld van een regionaal systeem is de stad Fuzhou waar het bedrijf JD Finance adverteert met het gebruik van AI om een 'smart city credit platform' te ontwikkelen.[885]

Er komt steeds meer informatie naar boven over de onderdrukking van islamitische Oeigoeren in de Chinese provincie Xingjiang. Veel aandacht gaat daarbij uit naar de 'heropvoedingskampen' waarin meer dan een miljoen mensen gevangen zitten. Relevant daarnaast is dat de *Strike Hard Campaign* die in 2014 is begonnen, ook een sterke digitale component heeft en dat daarbij technologieën als AI worden gebruikt. De vele politiecheckpunten in de provincie zijn uitgerust met biometrische sensoren, irisscanners en toegang tot CCTV-camera's in de buurt. Van veel Oeigoeren wordt regelmatig DNA verzameld en zij worden gedwongen de app Jingwang te installeren. Die maakt het niet alleen mogelijk om berichten te volgen en te blokkeren, maar geeft de autoriteiten ook direct toegang tot de telefoons van de gebruikers. De politie inspecteert of

883 Polyakova en Meserole 2019: 3-4.
884 Deze technologie wordt bijvoorbeeld gebruikt om chauffeurs te identificeren bij de transportapp Didi, om geld over te maken via Alipay, om treintickets op te halen en voor toegang tot toeristische attracties (Agrawal et al. 2018: 219).
885 Ahmed 2019: 57-59.

mensen deze 'elektronische handboeien' daadwerkelijk hebben geïnstalleerd.[886] Alle auto's in de provincie zijn verplicht om navigatiesoftware te installeren waar Beidou op zit, de Chinese versie van GPS, en op plekken waar de CCTV-camera's niet kunnen komen, worden zwermen drones ingezet. Naast de fysieke gevangenissenkent China, volgens een paper van het Brookings Instituut, dan ook "de grootste *open air* digitale gevangenis van de wereld".[887]

China toont dus bij uitstek hoe AI ingezet kan worden voor de doelen van autoritaire regimes. Het is des te belangrijker om hier zicht op te hebben, omdat de laatste jaren dergelijke technologieën ook meer en meer worden geëxporteerd. Verschillende soorten spelers doen dat: staatsorganen als het leger (de PLA) en het ministerie van Publieke Veiligheid, staatsbedrijven als CEIEC en private bedrijven als Huawei, ZTE en Tencent.[888]

Deze export gaat de hele wereld over.[889] De Chinese *Great Firewall* wordt opgezet in Vietnam en Thailand. Het bedrijf Yitu levert draagbare camera's met AI voor gezichtsherkenning aan de Maleisische politie en dong mee naar een gezichtsherkenningsproject in de publieke ruimte in Singapore.[890] Ethiopische veiligheidsdiensten gebruiken telecommunicatieproducten van ZTE om journalisten en activisten te monitoren. Zimbabwe en Angola hebben AI-deals getekend ten behoeve van hun regimes. In Venezuela heeft ZTE een contract voor een nationale ID-kaart, een betaalsysteem en een 'vaderland-database' waarmee het regime het Chinese socialekredietsysteem in het land kan invoeren. Surveillancesystemen worden gebruikt bij overheidscamera's in Ecuador en in Dubai wordt Chinese technologie gebruikt voor het programma Politie zonder politiemensen, waarin misdaad bestreden wordt met videosurveillance en gezichtsherkenningstechnologie.[891] Belangrijk om verder te noemen is dat Chinese bedrijven als Huawei en SenseTime samenwerkingen aangaan met universiteiten wereldwijd, ook in het westen.

Casus: Digitale dictatuur in Rusland

De andere grote speler in de ontwikkeling en export van digitale dictatuur is Rusland. De basis hiervoor ligt in een lange traditie van controle over informatie. Die gaat terug tot de Sovjet-Unie, maar kwam vrij snel na de val van dat

886 Singer en Brooking 2018: 101.
887 Polyakova en Meserole 2019: 5.
888 Weber 2019: 77-78.
889 De export van digitale autoritaire praktijken is vaak ook gewoon mensenwerk. Het zogenoemde '50 Cent Leger', waar twee miljoen Chinezen lid van zouden zijn, is vernoemd naar het bedrag dat ze zouden krijgen om positieve berichten over China te plaatsen.
890 Wright 2019a: 36.
891 Polyakova en Meserole 2019: 6.

autoritaire regime terug. In 1995 werd er al een wet aangenomen waardoor de FSB, de opvolger van de KGB, alle privécommunicatie van burgers kon monitoren. Sindsdien heeft een reeks van wetten de grip van de overheid op het RuNet, zoals het Russische internet wordt genoemd, vergroot. Deze wetten geven de mogelijkheid om websites te blokkeren, bloggers te registreren, data in het land op te slaan en de FSB toegang te geven tot encryptie. Druk op VKontakte, het grootste Russische socialemediaplatform, om onder andere toegang te geven tot informatie over de campagne van Alexei Navalny, bracht de baas van het bedrijf ertoe om zijn aandelen te verkopen. Later richtte hij de chatapplicatie Telegram op, die weer met de Russische autoriteiten zou botsen vanwege de encryptie die deze gebruikt.[892]

Net als we zagen bij China, wordt Ruslands instrumentarium van digitale dictatuur naar het buitenland geëxporteerd (zie ook tekstbox 8.6). Het belang van digitalisering bij conflicten wordt er al lang erkend. De Russische generaal Valery Gerimasov zou het gebruik van de asymmetrische mogelijkheden die het internet biedt voor internationale competitie benadrukt hebben. In het kielzog daarvan hebben 75 onderwijs- en onderzoeksinstellingen onder aansturing van de FSB de opdracht gekregen om te bestuderen hoe informatie tot wapen gemaakt kan worden. Een onderzoeker van de NAVO vat de strategie samen als de '4D's': *"dismiss the critic, distort the facts, distract from the main issue, and dismay the audience"*.[893] Traditionele en onlinemediakanalen zoals Russia Today, Sputnik en Baltica spelen een belangrijke rol in deze informatieoorlog. Ruslands inname van de Krim is ook wel 'Schrödingers oorlog' genoemd, vanwege de manier waarop het land desinformatie, verwarring en hybride oorlogsvoering gebruikte.[894]

Tekstbox 8.6 – De Russische toolbox

De instrumenten die Rusland gebruikt voor binnenlandse controle, zijn divers. Een belangrijk stuk hardware is de zogenoemde SORM. Dat is een doos die Internet Service Providers verplicht moeten installeren en waarmee de geheime diensten al het internetverkeer kunnen kopiëren en monitoren.[895]

Naast hardware berust digitale controle ook op mensenwerk. Betaalde trollenfabrieken, hacktivisten en het beruchte Internet Research Agency

892 Kerr 2019: 64-67.
893 Singer en Brooking 2018: 107.
894 Singer en Brooking 2018: 205.
895 Soldatov en Borogan 2015.

(IRA) in Sint-Petersburg worden gebruikt om online invloed te projecteren in binnen- en buitenland. Naast die instrumenten en methoden speelt AI ook een belangrijke rol in Ruslands digitale surveillance. Sinds 2015 is het land begonnen met het systeem Safe City, waarmee de herkenning van gezichten en bewegende objecten op de videobeelden van allerlei camera's direct met overheidsinstanties wordt gedeeld. Van 2012 tot 2019 investeerde het land 2,8 miljard dollar om alle steden van het Wereldkampioenschap voetbal in 2018 hiermee te voorzien. Meer dan 100.000 camera's die in Moskou geplaatst zijn, hebben gezichtsherkenningssoftware van het bedrijf NTechlab.[896]

Ook gebruiken de Russische veiligheidsdiensten het Semantic Archive Platform van het softwarebedrijf Analytical Business Solutions om open source-data te verzamelen, te verwerken en te analyseren.[897]

Er zijn duidelijke verschillen tussen de Chinese en de Russische exportproducten. Terwijl het Chinese 50 Cent Army er vooral op gericht is om positieve beelden over China te verspreiden, gaat het Rusland vooral om het verspreiden van negatieve berichten in landen waar het verdeeldheid wil zaaien. Nina Schick legt een verband tussen dit huidige beleid en de 'actieve maatregelen' ten tijde van de Sovjet-Unie. Het doel daarvan was om de perceptie van de werkelijkheid van anderen zo te veranderen dat ze niet meer in staat waren om zinnige conclusies te trekken over hoe hun eigen belangen verdedigd kunnen worden.[898]

Een tweede verschil is dat het Chinese exportproduct technologisch veel geavanceerder en duurder is, omdat daarmee het web goeddeels gecontroleerd wordt. Het Russische product daarentegen berust meer op bepaalde hardware, principes van intimidatie en wetgeving om de bevolking te sturen. Volgens een studie van het Brookings Institute zou het Russische product wel eens interessanter kunnen zijn voor landen die niet zo rijk zijn en weinig middelen hebben om het hele internet in hun land te controleren.

Het SORM-apparaat is wijdverspreid in landen uit de voormalige Sovjetsfeer, zoals Kirgizië, Wit-Rusland en Kazachstan. De bedrijven die het exporteren, Protei en Peter-Service, hebben ook telecombedrijven in het Midden-Oosten

896 Polyakova en Meserole 2019: 8.
897 Morgus 2019: 92.
898 Schick 2020: 54.

en Latijns-Amerika als klanten. Het *Semantic Archive Platform* is in gebruik in Wit-Rusland, Oekraïne en Kazachstan.[899]

Een laatste aspect van de export van de Russische digitale dictatuur betreft het beleid in internationale instituties en fora. Daarin poogt Rusland de grens tussen cybersecurity en informatiecontrole te laten vervagen, zodat het landen die om de eerste geven ook meegaan in de tweede. In de VN heeft het land documenten voorgelegd voor *een International Code of Conduct for Information Security*, waarmee mensenrechten en internationaal recht onder druk zouden komen te staan. Ook wil het land het internet onder de controle van de ITU brengen, waarmee het internet onder het bestuur van staten terecht zou komen. Ten slotte dringt het aan op internetbekabeling tussen de BRICS-landen – de pkomende economieën Brazilië, Rusland, India, China en Zuid-Afrika – die niet door de VS loopt.[900] We zijn in het eerste deel van dit hoofdstuk over competitiviteit het belang van dergelijke internationale fora tegenkomen. Genoemd Russisch beleid maakt duidelijk hoezeer onderhandelingen in die fora ook met veiligheidsvraagstukken verweven zijn.

Kernpunten – Digitale dictatuur in China en Rusland

- China en Rusland zijn de leiders op het gebied van digitale dictatuur en de export daarvan. Beide landen gebruiken een set van verschillende instrumenten voor autoritaire doeleinden. Die instrumenten worden naar alle uithoeken van de wereld geëxporteerd.
- Het Chinese model is technologisch hoogstaand en er vooral op gericht om positieve berichtgeving over het regime te stimuleren.
- Het Russische model is minder geavanceerd en gebruikt meer hardware en analoge vormen van intimidatie. In het buitenland is dit model vooral gericht op het scheppen van verwarring en conflict.
- Digitale dictatuur is een complex en veelzijdig fenomeen dat serieuze aandacht vereist.

8.3 Tot slot

De opgave van positionering gaat over de plek en de rol van een land in het internationale speelveld. Daarbij gaat het om de relatie tot andere landen, maar ook tot niet-statelijke actoren als bedrijven en criminele organisaties. In dit hoofdstuk hebben we gezien dat er bij onderwerpen als autonome wapens, regulering van AI en standaardisering verschillende internationale tonelen

899 Polyakova en Meserole 2019: 10.
900 Morgus 2019: 93.

zijn waarop geacteerd kan en moet worden, en ook welke uitdagingen daarbij komen kijken. Bij de beleidsimplicaties in het volgende deel gaan wij nader in op de vraag wat voor 'AI-diplomatie' dit van Nederland vergt.

Ook hebben we gezien hoe zorgwekkend de fenomenen als de informatieoorlog en digitale dictatuur zijn waarin AI een rol in speelt. Ze kunnen een bedreiging vormen voor vrijheid en democratie wereldwijd, maar ook in ons land. Dat vraagt inspanningen om grip te krijgen op deze fenomenen en er antwoorden op te kunnen formuleren.

De twee vraagstukken van concurrentiekracht en veiligheid zijn zeker in het kader van AI verweven. In dit hoofdstuk hebben we in dat verband gewezen op het belang van een AI-diplomatie en de bewustwording van de risico's van AI voor de nationale veiligheid.

Deel 3

Agenda: conclusies en aanbevelingen voor AI-beleid in Nederland

9. Beschouw AI als systeemtechnologie

AI is niet zomaar een technologie, maar een systeemtechnologie die onze samenleving fundamenteel zal veranderen. Dat is de hoofdboodschap van dit rapport. Overheid en samenleving dienen daarom veel bewuster en actiever met de inbedding van AI om te gaan. Voor de overheid geldt in het bijzonder dat zij gestructureerde aandacht moet hebben voor een vijftal opgaven als ze het proces van inbedding mede vorm wil kunnen geven. Alleen dan zal zij de publieke waarden rondom AI in de toekomst kunnen blijven beschermen. Een dergelijke uitdaging vraagt om een beleidsinfrastructuur waarin zowel politieke als ambtelijke inzet verankerd zijn.

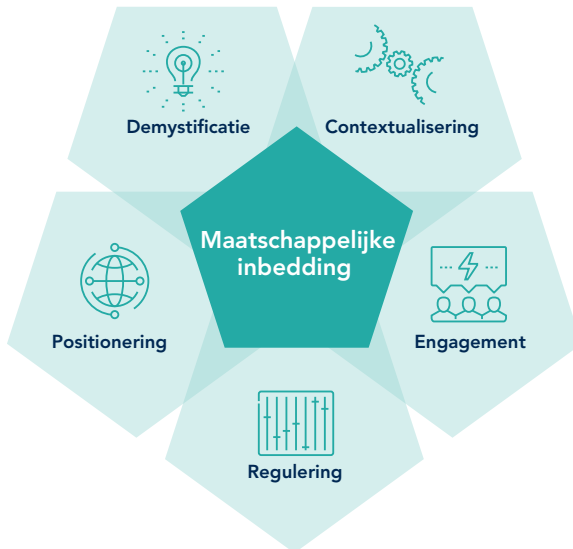
Kunstmatige intelligentie (AI) is in onze eeuw wat elektriciteit was in de negentiende of de verbrandingsmotor in de twintigste eeuw. Zij is geen concrete technologie die goed te overzien is en door een groep van experts of beleidsmedewerkers van een of enkele ministeries in goede banen valt te leiden. Omdat AI alomtegenwoordig is, continue verbetering kent en complementaire innovatie mogelijk maakt, is zij een veelzijdig en onvoorspelbaar fenomeen. Deze onvoorspelbaarheid van en daarmee onzekerheid over de wijze waarop de inbedding van AI de komende jaren vorm zal krijgen, mag en kan echter geen reden zijn om de ontwikkelingen op hun beloop te laten. De potentieel talloze publieke waarden die in het geding zijn, vragen om een grondig doordachte omgang met de inbedding van AI. De overheid moet zich bovendien rekenschap geven van de bredere agenda die met de inbedding van AI opdoemt, omdat naar de toekomst toe ook haar vermogen om te interveniëren en te corrigeren op het spel staat.

Juist door AI vanuit het brede perspectief van systeemtechnologieën te beschouwen, kunnen we volop leren van de inbedding van eerdere systeemtechnologieën. De lessen uit de inbedding van die eerdere systeemtechnologieën vormen in dit rapport het fundament voor de aanbevelingen die de WRR in dit slothoofdstuk presenteert. Kernpunt daarbij is dat het beschouwen van AI als een systeemtechnologie implicaties heeft voor de wijze waarop we naar publieke waarden kijken. Zoals we in hoofdstuk 1 hebben betoogd, leert de geschiedenis van systeemtechnologieën dat de impact ervan op de publieke waarden nooit met een lijst van deze waarden afgebakend kan worden. Immers, wanneer AI in potentie in de gehele samenleving toegepast kan worden en we pas aan het begin van die ontwikkeling staan, is het effect van AI op publieke waarden niet alleen breed, maar ook onvoorspelbaar. In de eerdere hoofdstukken en de slotanalyse in dit hoofdstuk benaderen we de publieke waarden daarom op een wijze die het dynamische karakter van de inbedding van AI in de samenleving recht doet.

9.1 Vijf opgaven als lessen uit het verleden

Op basis van een analyse van de geschiedenis van eerdere systeemtechnologieën onderscheidt de WRR vijf opgaven voor de overheid en de samenleving om richting te geven aan de inbedding van AI: **demystificatie** van wat AI is en kan, **contextualisering** van de ontwikkeling en toepassing, **engagement** van verschillende partijen, **regulering** van technologie, technologiegebruik en maatschappelijke implicaties en **positionering** ten opzichte van andere landen en internationale organisaties (figuur 9.1). Deze opgaven hebben wij in deel 2 van dit rapport uitgebreid besproken en lichten wij hier kort toe. We geven ook steeds aan welke publieke waarden daarbij in het geding zijn en benoemen wat het risico is als de opgaven niet ter hand worden genomen.

Figuur 9.1 Vijf opgaven voor de maatschappelijke inbedding van AI



AI als systeemtechnologie

Er is een rijke wetenschappelijke literatuur over technologische revoluties, epochale innovaties en technische tijdperken. Een centraal concept is dat van *'general purpose technologies'*: technologieën die niet voor een specifiek doel, maar breed door de hele samenleving toegepast kunnen worden. Eerdere voorbeelden hiervan zijn de stoommachine, elektriciteit, de verbrandingsmotor en de computer. We hebben in hoofdstuk 3 laten zien dat AI de drie eigenschappen van een *general purpose technology* bezit: het is (1) alomtegenwoordig, (2) kent

continue technische verbetering en (3) maakt complementaire innovaties in andere gebieden mogelijk.

In dit rapport duidt de WRR AI als een systeemtechnologie. Enerzijds wijzen we daarmee op het feit dat AI net als elektriciteit en verbrandingsmotoren onderdeel is van een breder systeem van andere technologieën. Anderzijds leggen wij met deze term de nadruk op het systemisch effect dat dergelijke technologieën op de samenleving hebben.

Wat verstaan wij onder AI?
In navolging van de High-Level Expert Group on AI (AI HLEG) van de Europese Commissie definiëren we AI in dit rapport als 'systemen die intelligent gedrag vertonen door hun omgeving te analyseren en acties te ondernemen – met enige graad van autonomie – om specifieke doelen te bereiken'.

De breedste definitie stelt AI gelijk met het gebruik van algoritmes en de strengste definitie ziet AI als de nabootsing van alle menselijke vaardigheden ('*artificial general intelligence*'). De eerste rekt het begrip van AI enorm op terwijl de tweede het juist weg definieert. De definitie van de AI HLEG is voldoende specifiek, maar laat – door niet-concrete technieken als *deep learning* te noemen – voldoende ruimte voor nieuwe technieken en ontwikkelingen.

AI en digitale technologie
AI is sterk verweven met andere digitale technologieën zoals de computer en data maar valt daar niet mee samen. Een van de vaders van de computer, Alan Turing, is eveneens de bedenker van de zogenoemde Turingtest die gebruikt wordt om AI-systemen te beoordelen. Grote hoeveelheden data, vooral internetdata, zijn ook cruciaal voor AI. Zeker de huidige methoden van *deep learning* vereisen grote hoeveelheden data om goed te kunnen werken. Tegelijkertijd valt AI niet tot die andere technologieën te reduceren. In hoofdstukken 1 en 2 schetsten we de ontwikkeling van AI, die op allerlei punten verbonden is met computers, data en het internet. Maar AI kent daarnaast een eigen wetenschap en geschiedenis met 'AI-lentes en -winters'. Terwijl computers al sinds de Tweede Wereldoorlog breed gebruikt worden en het internet vanaf de jaren negentig overal in de samenleving te vinden is, is de opkomst van AI als maatschappelijk fenomeen iets van de laatste jaren. Daarom verdient de technologie aparte aandacht.

Opgave 1: Demystificatie

De eerste opgave – *demystificatie* – gaat over de beelden die er over AI als technologie bestaan. In feite gaat deze opgave over de vraag: *Waar hebben we het over?* Rondom systeemtechnologieën ontstaan altijd extreme beelden. Te hooggespannen verwachtingen leiden tot desillusie en ondoordachte toepassingen, terwijl overtrokken angsten leiden tot afkeer en onbenutte mogelijkheden. Vooral op de langere termijn zal het vasthouden aan dergelijke beelden negatief uitwerken. De WRR stelt dat meer realisme nodig is om maatschappelijk, en wat betreft de publieke waarden, de juiste vragen te kunnen stellen. In het verleden ontstonden bij elektriciteit en auto's irreële verwachtingen door allerlei publieke tentoonstellingen en races. Daarnaast geloofden commentatoren dat treinen, telegrafen en het internet wereldwijde vrede zouden brengen doordat deze onderlinge verbindingen mogelijk maakten. Omgekeerd hebben beelden van Frankenstein en woorden als 'elektrocutie' – waardoor elektriciteit werd verbonden aan de dood – angsten gecreëerd rondom eerdere systeemtechnologieën.

Ook rondom AI bestaan er allerlei mythen. AI-systemen zouden rationeel en objectief zijn, maar ook werken als een onbegrijpelijke *black box*. De technologie zou alle menselijke vermogens kunnen evenaren en zelfs overstijgen, en zou zich zelfs tegen de mensheid kunnen keren. Daarnaast bestaan er allerlei mythen die de bredere digitalisering betreffen, zoals het tot recent populaire idee dat de ontwikkeling van het internet vrijgelaten en dus vooral niet gereguleerd moest worden. Maar ook dat er geen alternatief zou zijn voor de huidige vorm van digitale technologie en dat digitalisering een oplossing biedt voor ieder probleem.

Adresseren we dergelijke beelden niet, dan kan de samenleving te veel gaan vertrouwen op AI-systemen, met allerlei kwalijke gevolgen. Het kan ook leiden tot afkeer van AI, waardoor de vruchten ervan niet (voldoende) geplukt worden. Overtrokken beelden werken ten slotte in de hand dat cruciale vragen over de inbedding van de technologie niet publiek bediscussieerd worden. Met demystificatie zijn in het bijzonder publieke waarden gemoeid als de rechtsbescherming en het vertrouwen van burgers, adequate informatievoorziening en de kwaliteit van het publieke debat.

Opgave 2: Contextualisering

De tweede opgave die we onderscheiden, is *contextualisering*. Hierbij gaat het om de toepassing van AI en speelt de vraag: *Hoe gaat de technologie werken?* Dat betekent in de eerste plaats aandacht voor het technische ecosysteem. Systeemtechnologieën functioneren niet zelfstandig, ze zijn afhankelijk van andere, ondersteunende technologieën dan wel daarop gebaseerde faciliteiten. Denk aan de verbinding van de auto met de olie-industrie, tankstations en een wegennetwerk. Ook raken systeemtechnologieën met de tijd verbonden met

andere opkomende technologieën, zoals de auto verbonden is met elektronica. Naast het technische ecosysteem gaat contextualisering over het sociale ecosysteem. Op macroniveau is een langdurige inspanning nodig om werkprocessen, waardenketens en kennisontwikkeling aan te passen, waarna organisaties de technologie pas echt goed kunnen gebruiken en er productiever door worden. Op microniveau vraagt dit om gedragsverandering en een juiste interactie tussen gebruikers en de nieuwe technologie.

Ook AI vereist verschillende ondersteunende technologieën dan wel daarop gebaseerde faciliteiten, zoals data, telecommunicatienetwerken, chips en supercomputers. Bovendien zien we al een groeiende verbinding van AI met andere nieuwe technologieën, zoals 5G-netwerken, het Internet of Things en quantum computing. Wat betreft de ontwikkelingen op macroniveau lijkt de verwachting dat AI het mensenwerk massaal overbodig gaat maken, ongegrond. Eerder is een proces vereist van intensieve training en oefening om AI op de werkvloer effectief te kunnen maken. Op microniveau draait het om de vraag hoe een goede mens-machine-interactie tot stand kan komen, met de relatieve autonomie van AI-systemen als belangrijkste uitdaging.

Te weinig aandacht voor ondersteunende technologieën en faciliteiten als (kwalitatief) goede, veilige en voldoende beschikbare data en netwerken zal ertoe leiden dat AI-systemen slecht zullen functioneren, bepaalde kansen onvoldoende worden benut dan wel verdere ontwikkeling stagneert. Zoals het wegennetwerk essentieel was voor het gebruik van de auto, zo vereist AI technische aanpassingen aan de omgeving. Aandacht voor die omgeving is met name belangrijk voor gebieden waar de Nederlandse samenleving haar voordeel kan doen met AI. Dan gaat het zowel om gebieden waarop ons land zich van oudsher profileert (zoals de landbouw en dienstensector) als om gebieden waar AI een bijdrage kan leveren aan de aanpak van bestaande uitdagingen (zoals de zorg). Onvoldoende aandacht voor het sociale ecosysteem leidt eveneens tot gebrekkige implementatie, maar ook tot allerlei misstanden of afwijzing omdat gebruikers van AI-systemen onvoldoende geëquipeerd zijn. Op het spel staan dus de kwaliteit en de veiligheid van AI-toepassingen, maar ook de publieke baten die op allerlei terreinen te winnen zijn, variërend van een kwalitatief betere en breder toegankelijke zorg en onderwijs tot een betere dienstverlening door de overheid.

Opgave 3: Engagement

De derde opgave van *engagement* betreft de maatschappelijke omgeving van AI en gaat over de vraag: *Wie moeten er betrokken zijn?* Bij nieuwe systeemtechnologieën hebben grote bedrijven en overheden de middelen en belangen om vroege gebruikers van innovaties te zijn. Partijen in het maatschappelijk middenveld raken doorgaans pas later betrokken. Op die manier verscherpen deze nieuwe technologieën in eerste instantie bestaande maatschappelijke

machtsverhoudingen: de wijze waarop de stoommachine werd ingezet in fabrieksmatige productieprocessen, marginaliseerde de arbeiders. En bij het aanpassen van de infrastructuur voor de optimale introductie van de auto werden voorgangers – destijds voornamelijk armere burgers – van de weg verdreven.

Het engagement van belanghebbenden in de samenleving kan een breed spectrum aan vormen aannemen. Een extreme reactie is gewelddadig verzet. Maar ook protest en oproepen tot verboden zijn manieren om een nieuwe technologie aan banden te leggen. Aan het andere eind van het spectrum kan het maatschappelijk middenveld bijdragen aan de verbetering van de technologie, bijvoorbeeld door eigen expertise in te brengen of door die technologie in de eigen praktijk toe te gaan passen.

Ook AI draagt in sterke mate bij aan een verscherping van de bestaande machtsongelijkheden. Op allerlei manieren worden minder welvarende burgers, etnische minderheden en vrouwen door algoritmen benadeeld. Op een aantal thema's als autonome wapens, gezichtsherkenning en AI-gebruik door de politie mobiliseert het maatschappelijk middenveld zich inmiddels flink. Veel van deze mobilisatie neemt de vorm aan van protest en oproepen tot verboden, maar ook van werkstakingen. Als het gaat om meer coöperatieve vormen van engagement gericht op een betere inbedding van AI – zoals het inbrengen van de eigen expertise of het gebruik van AI bij uitdagingen op het terrein van klimaat, armoedebestrijding of mensenrechten – blijkt echter nog een wereld te winnen.

Wat als het engagement achterblijft? Dan is het aannemelijk dat bestaande machtsongelijkheden zich verdiepen en machtsrelaties tussen overheden en grote bedrijven enerzijds en burgers anderzijds verder scheefgroeien. Vooral de rechten van allerlei zwakkere maatschappelijke partijen worden dan bedreigd. Publieke waarden die op het spel staan als engagement achterblijft, zijn dus fundamentele rechten als gelijkheid, privacy, non-discriminatie en autonomie en democratische principes als inspraak, inclusie en pluriformiteit. Belangrijk voor het vormgeven van engagement is een regulerend kader, en daarmee komen we bij de vierde opgave voor de overheid.

Opgave 4: Regulering

Ten vierde onderscheiden wij *regulering*, een opgave die speelt op het niveau van de samenleving als geheel. De centrale vraag is hier: *Wat voor kaders zijn nodig?* Wanneer een nieuwe technologie het lab uit gaat, is het aanvankelijk moeilijk om de noodzakelijke kaders te overzien, aan te passen of te ontwikkelen. Veel is nog onduidelijk over de aard en de effecten van de technologie. Bovendien: zolang AI nog niet in de volle breedte en de talloze contexten van

onze samenleving is ingebed, valt moeilijk te overzien welke voor die contexten specifieke publieke waarden in het gedrang komen.

In de vroege fase benadrukken technologiebedrijven vaak dat zelfregulering door de sector adequaat is of dat gebruikers er zelf op moeten toezien dat bepaalde waarden geborgd zijn. Gaandeweg komen echter structurele vraagstukken aan het licht die een actievere rol van de overheid vereisen. Systeemtechnologieën kenden aanvankelijk bijvoorbeeld een machtsconcentratie bij enkele bedrijven, zoals GE en Westinghouse bij elektriciteit of de 'Big Three' uit Detroit bij auto's. Ook andere factoren dragen bij aan de noodzaak van een actievere overheidsrol. Naarmate de technologie dieper raakt ingebed in de samenleving, komen er ook meer en meer raakvlakken met de publieke waarden waar de overheid over waakt. Met de tijd ontstaat bovendien beter zicht op de bredere maatschappelijke effecten van een nieuwe technologie, wat beleid en wetgeving steeds minder tentatief maakt. Dat is het moment waarop de overheid een bredere wetgevingsagenda dient te ontwikkelen en niet langer kan volstaan met losse wetgevingsdossiers.

Ook bij AI zien we de aanvankelijke focus op zelfregulering. Inmiddels is het momentum verschoven naar een actiever overheidsingrijpen. De Europese concept-Verordening voor AI is daar een goed voorbeeld van. Ondertussen komen ook allerlei structurele vraagstukken aan het licht, die de overheid eveneens zal moeten aanpakken om greep te houden op de effecten van de technologie. Het gaat daarbij om de machtsconcentratie bij grote bedrijven, de groei van surveillance in de samenleving en de steeds grotere afhankelijkheid van de publieke sector van het bedrijfsleven.

Natuurlijk kent regulering van systeemtechnologieën geen panacees of 'zilveren kogels'. Een brede set aan maatregelen die over een lange periode tot ontwikkeling komt, is nodig om een technologie goed in de samenleving in te bedden. Neem de verbrandingsmotor die de auto mogelijk maakte: gordels, verzekeringen, nummerplaten, airbags, rijexamens, verkeersregels en verkeersborden hebben stap voor stap allemaal bijgedragen aan de maatschappelijke inbedding ervan – een proces dat tot op de dag van vandaag voortduurt omdat ook de auto en zijn omgeving in ontwikkeling blijven. Onmogelijk konden al die maatregelen destijds bij de introductie voorzien worden. Dat betekent echter niet dat de wetgever moet blijven dubben over wat de wenselijke aanpak zou kunnen zijn. De opgave van regulering vraagt zowel om een grotere rol van de Nederlandse overheid als om een bredere wetgevingsagenda. Wacht de overheid te lang met een dergelijke agenda, dan zal de dynamiek van de inbedding van AI de wetgever inhalen en hebben bepaalde spelers de macht inmiddels zo naar zich toegetrokken dat een weg terug nauwelijks nog mogelijk is. Op dat moment

verliezen de bestaande kaders aan geldingskracht en komt de inrichting van de samenleving op basis van publieke waarden onder druk te staan.

Opgave 5: Positionering

De laatste opgave die we onderscheiden, is *positionering*. Deze opgave heeft betrekking op het internationale toneel en betreft de vraag: *Hoe verhouden wij ons internationaal?* Het gaat hierbij ten eerste over de rol die een nieuwe systeemtechnologie kan spelen bij het stimuleren van het eigen verdienvermogen van een land. Historisch hebben nieuwe systeemtechnologieën als de stoommachine, elektriciteit en de verbrandingsmotor veel bijgedragen aan de competitiviteit van landen. Daarnaast hebben deze technologieën ook invloed op de aard en uitkomsten van internationale conflicten. De spoorwegen waren essentieel voor de overwinning van Pruisen op Frankrijk in 1871 en de codekrakers van de eerste computers droegen bij aan de overwinning op de Duitsers in de Tweede Wereldoorlog. Als gevolg van deze twee dynamieken ontstaat doorgaans het idee van een mondiale race rondom de nieuwe technologie en pogen sommige landen zelfs om die volledig binnen de eigen landsgrenzen te ontwikkelen en te behouden. De historie leert echter dat systeemtechnologieën een mondiaal karakter hebben en dat juist internationale samenwerking kan bijdragen aan het verdienvermogen en de veiligheid van landen.

Deze dynamiek speelt ook rondom AI. Er wordt veel gesproken over een 'AI-race', met de VS en China als koplopers. Veel landen hebben de afgelopen periode dan ook AI-strategieën ontwikkeld om aan die race mee te doen en het verdienvermogen met AI te versterken. Maar ook op het gebied van conflict en veiligheid groeit het besef van de impact van AI. De meest prominente toepassingen daarvan zijn de zogenoemde autonome wapens. Inmiddels zijn er verschillende internationale initiatieven om de ontwikkeling en omvang van dit nieuwe wapenarsenaal in goede banen te leiden. Tegelijkertijd is er een breed palet aan militaire en civiele toepassingen van AI die de veiligheid van burgers kunnen bedreigen.

Als Nederland zich onvoldoende positioneert op het gebied van AI en onvoldoende oog heeft voor bredere samenwerking op internationaal – met name Europees – niveau, zal het ten eerste kansen missen om met AI het verdienvermogen van ons land te versterken. Ten tweede betekent te weinig aandacht voor positionering dat Nederland onvoldoende zicht heeft en voorbereid zal zijn op de veiligheidsrisico's die met de technologie gemoeid zijn.

Vijf opgaven, vijf transities

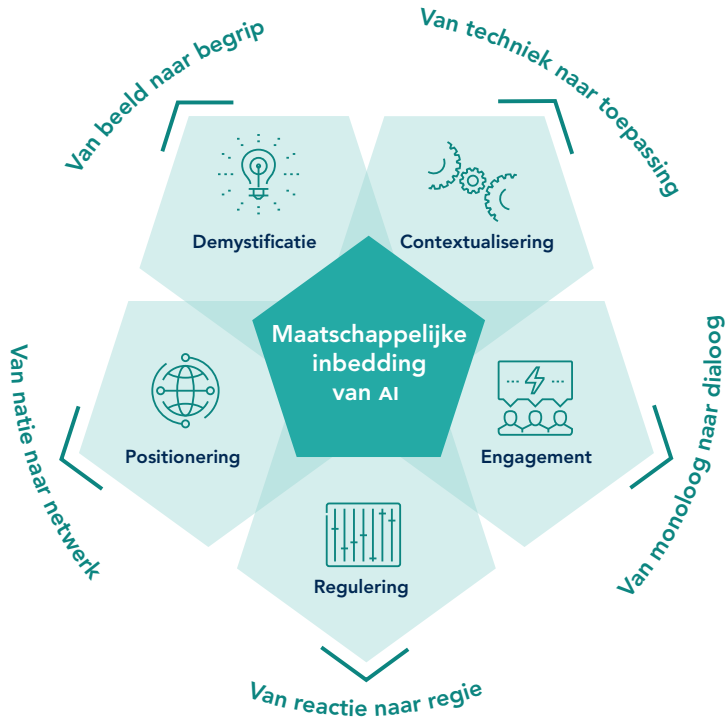
Deze vijf opgaven zijn kortom cruciaal bij het inbedden van AI in de samenleving. Het is belangrijk om te benadrukken dat ze met elkaar samenhangen. Demystificatie over AI versterkt bijvoorbeeld het vermogen van de samenleving

om zich met de technologie te engageren. Analytisch zijn ze dus te scheiden, maar in de praktijk is een integrale benadering nodig. Bij de inbedding van AI staat niet alleen veel op het spel (benutten innovatiepotentieel, maatschappelijke acceptatie, enzovoort), ook zijn daarbij telkens verschillende publieke waarden in het geding. Welke waarden dat zijn en wat het effect van AI daarop is, valt onmogelijk vooraf te voorspellen. We hebben in dit rapport betoogd dat het ondoenbaar is om een uitputtende lijst van publieke waarden op te stellen en deze in het licht van AI te analyseren. Het onvoorspelbare karakter van systeemtechnologieën noodzaakt tot een dynamischer perspectief. De WRR pleit er daarom voor om het debat over AI en de publieke waarden te voeren aan de hand van de vijf geïdentificeerde opgaven. Binnen het brede kader van deze opgaven kunnen behalve veel hedendaagse ook toekomstige vraagstukken een plek krijgen.

Met het cluster van de vijf opgaven biedt de WRR dus een langetermijnkader voor de omgang met AI. Maar daarmee is nog niet de vraag beantwoord wat er in het licht van die opgaven op dit moment dient te gebeuren, meer in het bijzonder vanuit de overheid. Met andere woorden, welke transities moeten er – gegeven de vijf opgaven – in gang gezet worden? Onderstaand formuleren we per opgave de te maken transitie (figuur 9.2), om die in de volgende paragraaf toe te lichten aan de hand van concrete aanbevelingen:

1. **Van beeld naar begrip**
2. **Van techniek naar toepassing**
3. **Van monoloog naar dialoog**
4. **Van reactie naar regie**
5. **Van natie naar netwerk**

Figuur 9.2 – Elke opgave vraagt om een transitie



Een brede agenda voor AI

De vijf transities representeren een AI-agenda voor de komende jaren. Een eerste opmerking die we daarbij maken, is dat de breedte van die agenda impliceert dat de overheid niet als enige verantwoordelijk is voor de invulling ervan. Allerlei partijen in de samenleving hebben bij de vijf opgaven een rol te spelen en verantwoordelijkheid te nemen. Zo zullen aan de transitie van beeld naar begrip wetenschappers moeten bijdragen. Maar ook burgers zelf kunnen deze transitie vormgeven door zich over AI te informeren of bijvoorbeeld de nationale AI-cursus te volgen. En media hebben een belangrijke rol richting burgers die niet zelf het initiatief (kunnen) nemen om zich nader te informeren. De transitie van techniek naar toepassing zal voor een groot deel door het bedrijfsleven uitgevoerd moeten worden. De overheid is een (potentieel) grote gebruiker, maar het zijn vooral ook allerlei private organisaties die de komende jaren moeten gaan uitvinden hoe AI in de praktijk goed kan functioneren. Kortom, bij alle opgaven en de bijbehorende transities is een collectieve inspanning van verschillende actoren nodig.

Een tweede opmerking is dat niet op alle onderdelen van de opgaven eenzelfde inspanning vereist is. Een aantal zaken zal vanzelf plaatsvinden. Zo kunnen we een zekere mate van demystificatie, en aldus een realistischer besef van de implicaties van AI, verwachten naarmate de samenleving daar collectief meer ervaring mee opdoet. Daarnaast komen er op sommige gebieden al initiatieven tot stand. Bij positionering bijvoorbeeld is er internationaal de nodige aandacht voor autonome wapens. De Nederlandse regering heeft daar een visie op ontwikkeld en gebruikt die in internationale organisaties als de VN. Of neem regulering. Niet elke nieuwe toepassing van AI vergt geheel nieuwe wetgeving. Bestaande regels bieden voor bepaalde toepassingen reeds de noodzakelijke kaders. Soms ook volstaat (vooralsnog) zelfregulering door bedrijven of andere partijen in de samenleving.

In dit rapport, dat gericht is aan de Nederlandse regering, maken we binnen de opgaven daarom een uitsnede: onze aanbevelingen betreffen zaken waarvan de WRR meent dat de overheid daar extra initiatief moet nemen. Bij elke aanbevelingen onderscheiden we daarom telkens een aantal concrete acties. Aan het slot van dit hoofdstuk geven we bovendien aan hoe deze aanbevelingen institutioneel en politiek ondersteund kunnen worden.

9.2 Transitie 1: Van beeld naar begrip

Bij de opgave van demystificatie gaat het om een transitie van 'beeld' naar 'besef'. Dat betekent dat de huidige aandacht voor de grote beelden – utopisch en dystopisch – plaats moet maken voor begrip. Het gaat, kortom, om meer afgewogen opvattingen over AI. De transitie die wij bepleiten, betekent niet dat de overheid de waarheid over AI aan de samenleving moet gaan verkondigen. Wel zal zij in het eigen handelen werk moeten maken van leren over AI, en dus over de volle breedte het evalueren van AI en de reflectie op de bedoelingen met AI tot een integraal onderdeel van haar functioneren moeten maken. Ook betekent het dat de overheid zich kritisch opstelt wanneer partijen met hoge verwachtingen over de mogelijkheden van AI spreken. Of andersom: als zij slechts de risicovolle scenario's schetsen. Onze eerste aanbeveling is dan ook gericht op het realiseren van deze transitie binnen de overheid zelf:

> Aanbeveling 1
> Maak leren over AI en de toepassing daarvan tot een expliciet doel bij het handelen door de overheid.

Bij het gebruik van een nieuwe technologie door de overheid zijn twee reflexen te onderscheiden. Enerzijds is er de reflex van 'technosolutionisme'. Een recent voorbeeld hiervan is de corona-app. Die werd al in een zeer vroeg stadium

aangekondigd als belangrijk onderdeel van het antwoord van de overheid op de COVID-19-crisis. Daarmee vond een vernauwing plaats van de discussie over dat antwoord. De vraag had eerst gesteld moeten worden wat de app aan dat antwoord zou kunnen bijdragen en of er op basis van kennis uit het veld – en de behoeften van bijvoorbeeld artsen en GGD's – andere, niet-technische oplossingen de voorkeur zouden moeten hebben. Er ging veel aandacht uit naar de ontwikkeling van de app. De termijn waarop die te realiseren was, bleek onderschat. Dat na de 'appathon' de conclusie moest zijn dat geen enkele kandidaat voldeed aan de voorwaarden voor de beoogde app, was zowel een teleurstelling voor degenen die er vertrouwen in hadden, als een bevestiging voor degenen die eerder wantrouwen hadden uitgesproken.

Anderzijds is er de reflex van technofobie, als gevolg van mislukte projecten of projecten die verboden werden. Zowel bij het uitkeren van toeslagen door de belastingdienst als bij het bestrijden van fraude met het Systeem Risico-indicatie (SyRI) door gemeenten, heeft onrechtmatig gebruik van bepaalde gegevens ernstige gevolgen gehad. Het gebruik van algoritmische systemen door de overheid en de potentiële consequenties daarvan zijn daarmee binnen een bepaald frame bij het grote publiek bekend geworden: de ondermijning van privacy en, in het verlengde daarvan, de schending van (fundamentele) rechten. Als gevolg daarvan heeft de overheid een angst ontwikkeld over het gebruik van algoritmen en AI in het bijzonder.[901]

Beide reflexen zijn niet vruchtbaar. Technosolutionisme leidt tot torenhoge verwachtingen, die na een mislukking omslaan in teleurstelling. De laatste tijd lijkt risicobewustzijn echter vooral technofobie in de hand te werken. Dat zorgt ervoor dat de overheid mogelijkheden mist om bestaande praktijken te verbeteren. Het is onvermijdelijk dat er fouten gemaakt zullen worden, maar dat mag er niet toe leiden dat de overheid afziet van de technologie als zodanig. Hoe hierin dan een goede balans te vinden?

Als systeemtechnologie vraagt AI om een lang proces van omgang, oefening en aanpassing. Het is geen simpel instrument of magische staf die kan worden ingekocht om het vervolgens zijn gang te laten gaan. Veel nadrukkelijker dan momenteel het geval is, moet leren daarom een expliciet doel worden van beleid. Dat betekent ook dat er (beleids)ruimte dient te zijn voor een mogelijke kans op fouten, binnen de marges gesteld door publieke waarden. Expliciet aandacht voor leren draagt dan ook bij aan het vermogen van bijvoorbeeld

901 Illustratief is het verbod om binnen de overheid discriminerende algoritmen in te zetten, dat eind mei 2021 met een ruime meerderheid in de Tweede Kamer werd aangenomen. AI kan immers ook voor positieve discriminatie worden benut.

uitvoeringsorganisaties om te experimenteren zonder dat zij er meteen politiek op afgerekend zullen worden. Om dit te realiseren is op directieniveau begrip en steun voor die experimenten noodzakelijk.

Leren en daarmee het opbouwen van capaciteiten rondom AI vergt wat betreft de WRR allereerst meer nadruk op het aantrekken van talent en scholing van personeel. AI zal een kernbestanddeel gaan vormen van de primaire processen van organisaties. Technisch en niet-technisch geschoolde medewerkers moeten daarvoor met elkaar het gesprek kunnen voeren en over en weer de juiste vragen kunnen stellen. Leren veronderstelt ook zorg voor basale zaken als tijdige en zorgvuldige archivering bij processen waar AI wordt ingezet, overdracht van kennis als mensen van baan veranderen en toegang tot databases en algoritmen. In verband met dat laatste wijst de WRR op de noodzaak om als overheid adequate contractuele afspraken te maken met private IT-leveranciers over de toegang tot en de beschikking over data, algoritmen en andere relevante AI-technieken. Ook hiertoe is het noodzakelijk dat de eigen kennis van de overheid van AI op peil is.

De genoemde zaken zijn niet slechts bijkomstigheden van grote AI-projecten, maar moeten expliciete doelen zijn die eraan voorafgaan. Dat heeft dus ook implicaties voor de werkwijze van de overheid. Vaak worden nu flinke begrotingen voor IT-projecten opgesteld met een vaste opleverdatum. Een meer iteratief proces, met kleinere projecten, valt te prefereren. Door te leren en te evalueren kunnen de juiste capaciteiten worden opgebouwd en is het vervolgens mogelijk op te schalen.

Overigens zijn niet alleen de uitvoerende instanties binnen de overheid gebaat bij een lerende aanpak gebaseerd op voortschrijdend inzicht in AI. Evenzeer geldt dit voor de politiek, de wetgever alsmede de toezichthoudende en rechtsprekende instanties. Een goed voorbeeld is de in hoofdstuk 7 behandelde de uitspraak van de Afdeling Bestuursrechtspraak Raad van State (ABRvS) in de zogenoemde AERIUS-zaak. De bestuursrechter formuleerde in deze zaak een toetsingskader voor vereiste transparantie. In de tweede uitspraak over de betreffende kwestie kwam de ABRvS tot een nadere precisering van de geformuleerde vereisten, mede ingegeven door hetgeen aan de hand van de eerste uitspraak was geleerd.

Concrete acties bij aanbeveling 1

- Maak werk van de opbouw van kennis, capaciteit en het voorkomen van afhankelijkheid.
- Schaal op vanuit kleinere ambities en projecten.
- Geef expliciet ruimte voor eventuele fouten en werk met korte cycli van evaluatie.

Ook de bredere samenleving is gebaat bij demystificatie, en dus bij een realistischer begrip van AI. Het gaat hierbij om meer dan kennis, namelijk ook om praktische vaardigheden en grip op hoe met AI om te gaan in verschillende contexten. Naar analogie van het begrip mediawijsheid zouden wij kunnen spreken van 'AI-wijsheid'. Voldoende AI-wijsheid in de samenleving is wezenlijk om mensen in staat te stellen op een realistische manier een omgang te vinden met AI en de veranderingen die daarmee gepaard gaan. Duidelijkheid over feitelijk AI-gebruik is hiervoor een belangrijke voorwaarde. Onze tweede aanbeveling is dan ook:

Aanbeveling 2
Stimuleer als overheid de ontwikkeling van AI-wijsheid bij het brede publiek, te beginnen met het opzetten van algoritmeregisters.

Allerlei partijen hebben een rol te spelen in het proces van demystificatie. Journalisten, wetenschappers en bedrijven kunnen allemaal bijdragen aan mythevorming of mythen juist ontkrachten. Demystificatie zal in de loop van de tijd dus deels vanzelf, zonder overheidssteun, op gang komen. Een zekere mate van AI-wijsheid op basis van basale kennis over AI kan burgers op termijn beter bestand maken tegen al te rooskleurige, zwartgallige of simpelweg onjuiste voorstellingen, hoewel deze waarschijnlijk nooit in hun geheel zijn uit te bannen. Op dit moment vindt echter een aantal ontwikkelingen plaats die de overheid ertoe zouden kunnen brengen een grotere rol te spelen.

Veel van de berichtgeving over AI is namelijk sensationeel. Speculaties over systemen die mensen voorbijstreven en op korte termijn de samenleving verstoren, komen veel voor. Er is behoefte aan meer feitelijke kennis over wat AI-systemen nu wel en niet kunnen. Veel van de berichtgeving voedt bovendien allerlei angsten, zoals beschreven in hoofdstuk 4. AI raakt ten slotte ook primair geassocieerd met toepassingen voor onontkoombare surveillance en controle, waardoor burgers de technologie kunnen gaan zien als een slechte ontwikkeling.

Een eerste stap die de overheid kan zetten om een realistischer beeld en meer begrip van AI te stimuleren, is door meer openheid te geven over het eigen gebruik van AI en wel door het opzetten van algoritmeregisters (zie tekstbox 9.1). De gemeente Amsterdam is daarmee begonnen om burgers inzicht te geven in waar en hoe zij algoritmen gebruikt. Utrecht en Rotterdam hebben dit initiatief inmiddels overgenomen. En in de voortgangsbrief 'AI en algoritmen' van 10 juni 2021 maakte het kabinet bekend, daartoe mede aangezet door de motie-Klaver, te onderzoeken "hoe een algoritmeregister bij kan dragen aan het

vergroten van transparantie over de inzet van algoritmen door de overheid".[902] In de 6 september 2021 aan de Tweede Kamer aangeboden I-strategie Rijk 2021-2025 is de opzet van een algoritmeregister een van de ambities waar de CIO's van alle ministeries en hun uitvoeringsorganisaties de komende jaren mee aan de slag gaan.[903]

Tekstbox 9.1 – Algoritmeregisters

Er gaan steeds vaker stemmen op voor de aanleg van algoritmeregisters. Op 19 januari 2021 nam de Tweede Kamer een motie aan met het voorstel om een algoritmeregister in te stellen. Het register moet gaan bijhouden welke algoritmen de overheid gebruikt, de doelen die zij daarmee nastreeft en welke data de algoritmen gebruiken (TK 2020-2021, 33510: 16). Aanleiding voor de motie vormde de parlementaire ondervraging naar aanleiding van de kindertoeslagaffaire.

Sommige steden, zoals Amsterdam en Helsinki, hebben al enkele algoritmeregisters gelanceerd. Het Amsterdamse register vermeldt onder meer algoritmen voor automatische parkeercontrole, de afwikkeling van meldingen over de openbare ruimte, de handhaving van illegale woningverhuur, en het monitoren van mensenmassa's. Per algoritme toont het register welke data is gebruikt om het te trainen, hoe het algoritme wordt ingezet, hoe mensen de voorspelling ervan gebruiken en welke afwegingen er over vervormingen (bias) of risico's zijn gemaakt.

De discussie over algoritmeregisters wordt ook internationaal gevoerd, zoals blijkt uit een rapport van de Law Society of England and Wales, de Britse beroepsorganisatie voor juristen. Deze organisatie bepleit een register voor algoritmen in het strafrecht, waarin voor elk algoritme belangrijke kenmerken, zoals transparantie, de standaardwerking en de gehanteerde data, worden vastgelegd. Ook de Europese Commissie lijkt te preluderen op een algoritmeregister: het voorstel voor de Verordening voor AI stelt de registratie van 'riskante' AI-systemen verplicht, ook voor privaat gebruik.

Burgers inzicht geven in de verschillende soorten toepassingen van AI die binnen de overheid verkend of gebruikt worden, is een noodzakelijke stap. De WRR wijst er echter op dat de instelling van algoritmeregisters pas echt meerwaarde

902 Zie: Kamerstukken II 2020-21, 26 643, nr. 765, actie 5.
903 Zie: Kamerstukken II, 2020-21, 26643, nr. 779, thema 6 speerpunt 2.

heeft als daarmee ook het gesprek over AI-gebruik in gang wordt gezet – zowel bij degenen die de toepassingen gebruiken als bij degenen die ermee te maken krijgen. Of dit daadwerkelijk gebeurt, hangt sterk af van de kwaliteit van de geboden informatie, het vermogen van de samenleving hiermee aan de slag te gaan en de reactie van de verantwoordelijke partijen op geconstateerde problemen. De instelling van algoritmeregisters moet daarom gepaard gaan met de verplichting ze periodiek te evalueren.

Daarnaast adviseert de WRR de overheid om alert te zijn op haar eigen rol bij het ontstaan van beelden in de samenleving over AI. De manier waarop de overheid AI inzet en daarover communiceert, kan immers bepaalde beelden en emoties oproepen. Om te beginnen is het noodzakelijk dat de overheid meer investeert in AI-toepassingen die de samenleving ten goede veranderen of een bijdrage leveren aan grote opgaven als klimaatverandering en bestrijding van sociale ongelijkheid. In dit rapport hebben we veel potentieel positieve voorbeelden besproken van de inzet van AI bij het gebruik van diagnostiek, gezondere lucht, minder energieverbruik, hulpverlening en dierenwelzijn. De WRR roept de overheid op om meer van dit type AI-toepassingen te stimuleren en te faciliteren alsmede de verdiensten daarvan onder de aandacht te brengen.

Vervolgens verdient ook de naamgeving van projecten meer aandacht. Dat SyRI in eerste instantie de naam 'Black Box' kreeg, kan onbedoeld hebben bijgedragen aan het beeld van AI als iets waar de mens geen betekenisvol inzicht in kan hebben. Maar denk ook aan het gebruik van termen als *'killer robots'* voor autonome wapensystemen of de 'robotrechter' voor AI binnen het recht. Zulke aanduidingen roepen direct sterke associaties op. Soms is dat de bedoeling, bijvoorbeeld om een duidelijk beeld te scheppen of een discussie op scherp te stellen. Sterke termen kunnen echter ook afleiden van de kwesties die er werkelijk toe doen. De kracht van de woorden die de overheid gebruikt, moet dus niet worden onderschat.

Met voorlichtingscampagnes en educatieve programma's kan de overheid bijdragen aan de ontwikkeling van basiskennis en vertrouwdheid met AI onder het brede publiek en mensen bewust maken van mogelijke valkuilen. In het Strategisch Actieplan voor AI (SAPAI) onderkent zij het belang van kennis over AI en het ontwikkelen van de juiste vaardigheden om ermee om te gaan. Daarmee lijkt er met name aandacht te zijn voor respectievelijk het onderwijs en de beroepsbevolking. Dat kan suggereren dat AI iets is voor de toekomst of waarmee mensen beroepsmatig te maken krijgen, terwijl het – vanwege de potentieel brede inzetbaarheid van AI – voor de hele samenleving van belang is om een basaal begrip te hebben van deze technologie. Allereerst is het van belang om burgers in staat te stellen alert te zijn op het enkele feit dat AI wordt toegepast. Het genoemde algoritmeregister, maar vooral ook het door de WRR

bepleitte gesprek over AI-gebruik, heeft bovendien alleen dan meerwaarde als de hele samenleving over een basaal begrip van de technologie beschikt. Op voorwaarde van een realistische voorlichting en educatie kan een beter begrip bij burgers ook het vertrouwen in AI vergroten (zie ook tekstbox 9.2). Bijzondere aandacht hierbij verdient de zoektocht naar effectieve manieren om dit basale begrip van AI ook te kunnen bieden aan burgers die minder zelfredzaam zijn.

Tekstbox 9.2 – Lopende trajecten voor AI-wijsheid

Voor geïnteresseerde burgers zijn er al verschillende initiatieven. De Nationale AI-cursus is in dit kader een aan te moedigen initiatief, evenals de door de NLAIC en TU Delft gelanceerde Nederlandse versie van de cursus *Elements of AI*. Degenen die wel zelf op zoek gaan naar informatie over de inzet van AI, zouden kunnen worden bediend door centrale informatielocaties, zoals de Kennisbank van het ministerie van Binnenlandse Zaken (BZK). In aanvulling daarop zou kunnen worden nagedacht over een digitale overzichtspagina met informatie over zaken waarop mensen moeten letten als ze AI willen inzetten of als ermee te maken krijgen, vergelijkbaar met bestaande pagina's over bijvoorbeeld het kopen van een huis of de bescherming van consumenten.

Meer nog dan om technische kennis, gaat het bij het ontwikkelen van AI-wijsheid om de kennis die nodig is om de berichtgeving over AI en de toepassingen van de technologie in perspectief te kunnen plaatsen en gevoel te krijgen voor de mogelijkheden en beperkingen ervan. Net zoals mediawijsheid verwijst naar de competenties die nodig zijn om deel te kunnen nemen aan de samenleving waarin media een belangrijke rol spelen.

Concrete acties bij aanbeveling 2

- Richt binnen de overheid een algoritmeregister voor AI-toepassingen op, zet daarmee het gesprek over AI-gebruik in gang en zorg voor periodieke evaluatie.
- Evalueer als overheid kritisch de eigen bijdrage aan de beeldvorming.
- Geef meer prioriteit aan (het faciliteren van) AI-toepassingen die ten goede komen aan de samenleving en geef daar ruchtbaarheid aan.
- Draag verder bij aan voorlichting en educatieve programma's gericht op de hele samenleving.

9.3 Transitie 2: Van techniek naar toepassing

De voor de opgave van contextualisering noodzakelijke transitie beweegt van techniek naar toepassing. Er is tegenwoordig veel aandacht voor de kenmerken van AI-systemen en bijbehorende vragen over transparantie en uitlegbaarheid. De opgave om de aandacht te verbreden naar de toepassing van AI betekent meer oog hebben voor de contexten waarin de technologie gebruikt wordt. Met name rondom de technische vereisten van die omgeving en de manier waarop gebruikers omgaan met AI.

Allereerst richten we ons voor deze transitie op het bredere technische ecosysteem waarin AI wordt toegepast. De aanbeveling op dit punt luidt:

> Aanbeveling 3
> Kies expliciet voor een Nederlandse AI-identiteit en onderzoek waar in de betreffende domeinen aanpassingen aan de technische omgeving nodig zijn.

Onze ecosysteembenadering laat zien hoeveel technologie nodig is voor een goedwerkende AI. Permanente aandacht is nodig voor het talent en het onderzoek voor algoritmeontwikkeling, de kwaliteit van netwerken, de toegang tot chips, de opbouw van databases, het ontwikkelen van cross-sectorale standaarden en het vertrouwd kunnen delen van data(sets). Ook is het belangrijk zicht te hebben op emergente technologieën die AI een impuls kunnen geven. Op diverse van deze terreinen heeft de overheid inmiddels initiatieven in gang gezet, onder andere via stimulering binnen het Groeifonds en de 'interbestuurlijke datastrategie'.[904]

Het specifieke aandachtspunt dat de WRR aanvullend meegeeft, betreft technische aanpassingen aan de omgeving waarin AI moet functioneren. We hebben daarvoor het woord '*enveloping*' gebruikt: het aanpassen van de omgeving zodat een technologie er goed in kan werken, zoals eerder de aanleg van het wegennet voor de auto of de energietoevoer voor elektrische apparaten. Die aanpassingen aan de omgeving kunnen vaak niet goed alleen door marktpartijen worden gedaan. Bovendien kunnen de gemaakte keuzes verstrekkende maatschappelijke gevolgen hebben, waardoor ze per definitie actieve betrokkenheid van de overheid vergen. Ging het bij de auto in de twintigste eeuw om het begaanbaar maken van de omgeving, bij AI gaat *enveloping* om het leesbaar maken van die

904 Kamerstukken II 2020-2021, 26 643, nr. 765.

omgeving: AI-systemen moeten die omgeving goed kunnen analyseren om er intelligent mee te kunnen interacteren (zie ook tekstbox 9.3).

Tekstbox 9.3 – Voorbeelden van *enveloping*

Neem de zelfrijdende auto. Die bevat steeds meer intelligentie, maar is nog ver verwijderd van het autonoom kunnen bewegen in een complexe omgeving. Aanpassingen in het wegdek, de belijning of zelfs de aanleg van specifieke infrastructuur waar ander verkeer niet komt – zoals de snelweg in het geval van de klassieke auto – kunnen grote stappen betekenen voor het gebruik. We hoeven niet meteen te denken aan nieuwe banen op de weg, maar bijvoorbeeld wel aan aanpassingen aan borden en signalen en speciale gebieden als industrieterreinen of andere gecontroleerde omgevingen waar veilig geëxperimenteerd kan worden.

Hetzelfde geldt voor allerlei 'domotica' of complexe industriële robots. Die kunnen nog maar slecht inspelen op de complexiteit en onvoorspelbaarheid van menselijk gedrag. Experimenten en investeringen in beter leesbare omgevingen kunnen een belangrijke bijdrage leveren aan het functioneren en daarmee het gebruik van die technologieën.

Voor de overheid is het echter ondoenlijk om de helpende hand te bieden bij het leesbaar maken van alle domeinen waarin AI-toepassingen hun weg zouden kunnen vinden. De overheid zal zich wat betreft de stimulering van AI dus noodzakelijkerwijs op een aantal specifieke domeinen moeten richten. De WRR bepleit daarom dat een beeld wordt ontwikkeld van wat we de Nederlandse 'AI-identiteit' noemen. Deze AI-identiteit bestaat uit de domeinen waarop ons land AI vooral wil ontwikkelen en inzetten. Domeinen waarvan de overheid vindt dat ons land het risico niet kan lopen dat de noodzakelijke aanpassingen van de AI-omgeving niet van de grond komen, vanwege coördinatieproblemen dan wel andere redenen waarom marktpartijen een en ander niet alleen kunnen of dienen te realiseren.

Tot de Nederlandse AI-identiteit kunnen de sectoren behoren waarin Nederland traditioneel sterk is dan wel die in belangrijke mate de motor van onze economie vormen, zoals onderdelen van de land- en tuinbouw, de infrastructuur en de logistiek. Het ontwikkelen van AI op deze domeinen zorgt ervoor dat Nederlandse bedrijven niet het risico lopen marktaandeel te verliezen of te afhankelijk worden van toeleveranciers, en creëert tegelijkertijd nieuwe verdienmodellen. De topsectoren zouden hieraan richting kunnen geven. Daarnaast, zo merkt de WRR op, kan de Nederlandse AI-identiteit ook domeinen omvatten die maatschappelijk van groot belang zijn en de overheid een

specifieke verantwoordelijkheid heeft om een voortrekkersrol te vervullen, zoals de gezondheidszorg of een goed functionerende overheid. De Nederlandse AI-Coalitie ontwikkelt al plannen om AI-innovatie in verschillende sectoren te stimuleren. Door het formuleren van een AI-identiteit zou de Nederlandse overheid aan dit proces meer richting kunnen geven. Daarmee kunnen ook knelpunten in de geïdentificeerde domeinen gericht worden aangepakt, bijvoorbeeld dat in delen van de landbouw de gebruikte modellen, analysetools, algoritmen en informatiediensten vaak in handen zijn van een beperkt aantal aanbieders of de onzekerheid over het eigendom van en de beschikkingsmacht over gezondheids- en zorgdata, zoals in de zorg.[905]

De overheid kan de Nederlandse AI-identiteit ook ondersteunen door strategisch gebruik te maken van het aanbestedingsbeleid. De overheid is een grote economische actor en kan markten stimuleren door de vraag naar bepaalde producten te ontwikkelen. De organisatie PIANOo houdt zich bezig met innovatiegericht inkoopbeleid en in 2019 heeft de overheid met SBIR (Small Business Innovation Research) een oproep gedaan aan bedrijven om innovatieve AI-toepassingen in de publieke sector te ontwikkelen. De overheid kan dit instrument intensiever gebruiken. Op allerlei gebieden van onderwijs tot lokale overheden is de inkoop bovendien versnipperd. Door meer en gerichter gebruik te maken van inkoopinstrumenten en inkoopeisen en standaarden te coördineren, kan de overheid de AI-ontwikkeling versterken en meer richten op voor Nederland belangrijke toepassingsgebieden.

Concrete acties bij aanbeveling 3

- Definieer de domeinen en aandachtsgebieden van een Nederlandse AI-identiteit.
- Breng technische benodigdheden en keuzes in die gebieden in kaart.
- Geef vorm aan de AI-identiteit door het aanbestedingsbeleid hierop aan te passen.

De transitie van techniek naar toepassing gaat niet alleen over de technologische context van AI, maar ook over de gedragscontext ervan en de mens die ermee aan de slag gaat. Onze aanbeveling voor dit sociale ecosysteem betreft daarom de volgende:

905 Vergelijk TNO 2021.

Aanbeveling 4
Versterk de vaardigheden en het kritische vermogen van individuen die met AI-systemen werken en ontwikkel daarvoor een stelsel van opleiding en certificering.

Volgens de WRR dient er meer aandacht uit te gaan naar mens-machine-interactie. Zelfs als technische systemen naar behoren werken en voldoen aan ethische richtlijnen, kan er in de praktijk nog veel misgaan. Bijvoorbeeld wanneer mensen niet goed weten hoe zij die AI-systemen moeten hanteren, of het functioneren daarvan niet kritisch weten te beoordelen. Een belangrijk gegeven is dat AI bestaande werkpraktijken transformeert, waardoor de rol van de menselijke gebruiker verandert en klassieke waarborgen tekort kunnen schieten. Zo kunnen we bijvoorbeeld wel eisen dat de mens altijd een beslissing neemt (*in* of *on the loop* blijft[906]), maar een cruciale vraag daarbij is hoe die verantwoordelijkheid ook betekenisvol en realistisch blijft.

Van een automobilist kan gezien de menselijke reactietijd bijvoorbeeld niet verwacht worden dat hij ingrijpt bij een zelfrijdende auto op het moment dat een ongeluk dreigt. Dat geldt ook voor mensen die steeds complexer wordende analysemethoden moeten overzien, interpreteren en controleren. Mensen die in hun werk ervaren dat algoritmen het vaak bij het goede eind hebben, zijn bovendien steeds minder geneigd om de uitkomsten ervan in twijfel te trekken (*automation bias*), zeker als de werkdruk hoog is. Nominaal is een mens nog verantwoordelijk en de menselijke factor dus nog aanwezig, maar die rol strookt dan niet meer met de werkelijke praktijk.

Een specifiek vraagstuk van mens-machine-interactie is hoe om te gaan met de feilbaarheid van beide. Wanneer mens en machine tot andere uitkomsten komen, kan het moeilijk zijn te beoordelen wie van de twee het bij het juiste eind heeft. Met menselijke kennis kan een fout van een algoritme in sommige situaties worden gecorrigeerd, maar een algoritme kan ook patronen ontdekken die een mens zelf niet had bedacht of verwacht. Hoe kunnen we het gebruik van AI zo organiseren dat er ruimte is voor de mens om de machine te corrigeren, en andersom?

906 Een mens *in the loop* betekent dat een AI-systeem alleen kan functioneren als een mens een bepaalde handeling verricht. Bij *on the loop* kan het systeem wel zelfstandig functioneren, maar een mens kan daarbij ingrijpen.

Tekstbox 9.4 – Het uitbreiden van denkkracht

Er zijn algoritmes die de politie adviseren om in een bepaalde buurt te patrouilleren of die een docent een schooladvies geven over een leerling. In beide gevallen kan het algoritme een fout maken. In het geval van de politie is het bijvoorbeeld voorgekomen dat het algoritme van het Criminaliteits Anticipatie Systeem (CAS) agenten naar parken stuurde om autodiefstal tegen te gaan. De redenering daarachter was dat autodiefstal bij samenscholingen plaatsvindt en dat die in het park te vinden zijn. In parken mogen echter geen auto's komen, dus iemand met gezond verstand zal dit advies afwijzen. Anderzijds is het in andere gevallen heel goed mogelijk dat het algoritme een patroon van criminaliteit heeft ontdekt waar mensen nog niet aan gedacht hebben.

Bij onderwijsinstellingen is het goed als docenten de uitkomsten van algoritmen niet zomaar naast zich neerleggen. Ze moeten er echter ook niet blind op vertrouwen (*automation bias*).

Het gaat er dus om contexten te creëren waarin de docent of de agent ondersteund wordt in zijn of haar werkzaamheden op een manier die rekening houdt met de feilbaarheid van zowel mens als machine. Anders geformuleerd, in plaats van menselijke intelligentie te vervangen, moet AI die vergroten en versterken ('*augmented intelligence*' of '*hybrid intelligence*').

Ook het interpreteren van AI-systemen is een vraagstuk van mens-machine-interactie (zie ook tekstbox 9.4). Dat betekent bijvoorbeeld dat gebruikers weten wat de aard is van de informatie die het AI-systeem genereert. Dit vraagt om kennis van het verschil tussen correlatie en causaliteit, van hoe hoog de foutmarge is en of een specifiek algoritme meer foutpositieven of meer foutnegatieven genereert. Een goede omgang betekent dat gebruikers weten wat de mogelijkheden en wat de grenzen zijn van de systemen waar zij mee werken.

Verschillende actoren kunnen hieraan bijdragen. In het bijzonder kunnen de verschillende AI-labs die in Nederland zijn opgericht, hier een belangrijke rol bij spelen. Naast zelf deel te nemen aan labs, zoals gebeurt in het Politielab, kan de overheid hier ook een stimulerende rol spelen. Ten eerste door in het eigen gebruik van AI meer aandacht te hebben voor de dynamiek van de mens-machine-interactie. Daarnaast dient zij bij interne processen van auditing, toezicht en het hanteren van richtlijnen meer aandacht te geven aan de gedragscontext en de eisen die gelden voor het gebruik van de technologie.

Wat daarnaast nodig is om een goede mens-machine-interactie, en dus ook de vaardigheden van mensen die met AI werken, te versterken, is het opzetten van een stelsel van opleiding en certificering voor zowel mens als machine. Hierbij valt te denken aan keurmerken, licenties en vereisten bij bepaalde toepassingen van AI. Voor dat laatste biedt de Europese concept-Verordening voor AI, en de daarin onderscheiden risiconiveaus, een goede grondslag. Bij licenties kunnen we een analogie trekken met hoe nieuwe medicijnen door industriële massaproductie op de markt worden gebracht. Hiervoor hebben publieke instanties een systeem van licenties en goedkeuring opgezet dat de toegang van medicijnen tot de markt bewaakt. Daarnaast zijn bijsluiters verplicht gesteld, zodat gebruikers van medicijnen zelf kennis kunnen nemen van bijwerkingen en mogelijke risico's. Keurmerken zijn er op allerlei domeinen van duurzame voedselproductie tot het voldoen aan standaarden voor het gebruik van chemische middelen, zoals de Europese REACH-wetgeving (*Registration, Evaluation, Authorisation and Restriction of Chemicals*). Organisaties die aan bepaalde standaarden voldoen, krijgen een keurmerk.

Een goede mens-machine-interactie verlangt behalve certificering die aanhaakt op het niveau van een product of organisatie, ook certificering op het niveau van het individu. In verschillende domeinen kennen we certificering van mensen die bepaalde technologieën gebruiken of bepaalde verantwoordelijkheden hebben. Denk aan de elektricien die de bedrading van een gebouw mag veranderen, de BIG-registratie in de gezondheidszorg en andere certificaten om bepaalde beroepshandelingen te mogen uitvoeren, zoals bij registeraccountants. Daarnaast kennen tal van functies, ook binnen de publieke sector, vereisten voor al dan niet permanente educatie. In feite hebben we het hier over verschillende vormen van een vaardigheidsbewijs of brevet.

De WRR bepleit niet dat iedereen die met AI te maken krijgt, geschoold dient te zijn en over een vaardigheidsbewijs of brevet zou moeten beschikken. Bij de systeemtechnologie elektriciteit hebben alle burgers ermee te maken, maar alleen de monteur die speciale verantwoordelijkheden heeft dient gekwalificeerd te zijn. Ook voor de inzet van AI geldt dat vrijwel iedereen hiermee te maken zal krijgen, maar dat alleen degenen die deze technologie hanteren dan wel voor de inzet ervan verantwoordelijk zijn, de daartoe benodigde kennis en vaardigheden moeten kunnen aantonen. De WRR benadrukt dat het daarbij niet alleen om een voldoende technisch begrip gaat, maar ook om het vermogen om vast te kunnen stellen of aan gestelde waarborgen is voldaan.

Concrete acties bij aanbeveling 4

- Besteed bij auditing, toezicht en het hanteren van richtlijnen expliciet aandacht aan de gedragscontext en mens-machine-interactie.
- Ontwikkel naast keurmerken, licenties en risiconiveaus gericht op AI-systemen en organisaties ook maatregelen voor de noodzakelijke kennis en kunde van mensen die met AI omgaan, zoals een vaardigheidsbewijs of brevet voor AI (zie tekstbox 9.5).

Tekstbox 9.5 – AI-brevetten

Hoe AI-brevetten eruit kunnen zien, voor wie ze bedoeld zijn en of ze verplicht moet worden gesteld, dient nader onderzocht te worden. Wij geven een aantal overwegingen mee:

- Kijk naar andere vaardigheidsbewijzen zoals de BIG-registratie, een vliegbrevet en de diplomering van monteurs voor wat toepasselijk is op het domein van AI.
- Wie precies een vaardigheidsbewijs dient te behalen – de ontwikkelaar, de instantie die AI inzet of de individuele eindgebruiker – zal variëren per gebruikscontext. AI in de vorm van een zorgrobot vraagt een andere benadering dan AI in zuiver algoritmische vorm.
- De scholing voor brevetten dient een theoretische component te bevatten. De kern ervan dient echter de praktijk te zijn: hoe moet gehandeld worden, wat zijn de mogelijkheden van de technologie, op welke signalen moet gelet worden, hoe vast te stellen of aan de waarborgen is voldaan en er moet vooral veel oefening zijn. Een duikbrevet vraagt theoretische kennis maar vooral veel oefening, wil iemand veilig de diepte in kunnen gaan. Brevetten strekken dus aanzienlijk verder dan de bestaande AI-cursussen, die het algemene begrip en de basistheorie betreffen.
- De praktische kennis bevat ook een set met handelingen die verricht moeten worden in complexe situaties of bij problemen. Denk aan medische standaarden voor specifieke handelingen binnen de gezondheidszorg. Ook moeten gebruikers van AI-systemen weten wanneer zij problemen zelf kunnen en mogen oplossen dan wel hulp van deskundigen moeten inschakelen.
- Gezien de enorme dynamiek op het gebied van AI is het raadzaam om brevetten te koppelen aan een vorm van permanente educatie.

9.4 Transitie 3: Van monoloog naar dialoog

Bij engagement is een transitie nodig van monoloog naar dialoog. Met monoloog doelen we op de situatie dat vraagstukken rondom AI nu goeddeels door een groep technische specialisten worden ingebracht, terwijl dit ook een zaak zou moeten zijn van allerlei andere partijen en organisaties. De term monoloog verwijst ook naar de grote afstand tussen de ontwikkelaars van AI-systemen en de maatschappelijke omgeving waarin die systemen worden toegepast. Burgers en partijen in het maatschappelijk middenveld hebben expertise in te brengen, maar hebben ook een belangrijke rol bij het geven van terugkoppeling over de werking van AI-systemen. Het gesprek over de vormgeving en toepassing van AI vergt kortom een grotere diversiteit aan partijen. Het kabinet onderneemt reeds initiatieven om het maatschappelijk middenveld te betrekken bij de ontwikkeling en toepassing van op AI gebaseerde applicaties. Illustratief is het voornemen van het kabinet om "het bedrijfsleven in samenwerking met consumentenorganisaties te bewegen tot het opstellen van een gedragscode op het gebied van online keuzebeïnvloeding door gebruik van consumentendata en algoritmes."[907] Willen consumentenorganisaties en andere organisaties die de rechten en belangen van burgers behartigen, hun rol echter kunnen waarmaken, dan moeten zij ook over de capaciteiten daarvoor kunnen beschikken. Om van monoloog naar dialoog te komen luidt onze eerste aanbeveling aan de overheid daarom:

> Aanbeveling 5
> Versterk de capaciteit van maatschappelijke organisaties om hun werk te verbreden naar het digitale domein, in het bijzonder met betrekking tot AI.

In het maatschappelijk middenveld bevinden zich verschillende partijen die de vraagstukken rondom AI goed op het vizier hebben. Dat geldt ten eerste natuurlijk voor organisaties die zich expliciet bezighouden met het digitale domein, zoals Waag, Bits of Freedom en Privacy First. Zij slagen er steeds beter in om het brede publiek te bereiken en kwesties rondom AI te agenderen. Ook grote mensenrechtenorganisaties als Amnesty hebben inmiddels ruim aandacht voor de impact van AI. Dat geldt echter in veel mindere mate voor organisaties die zich richten op bijvoorbeeld de belangen van werknemers, patiënten, leraren, mensen in armoede en achtergestelde en gediscrimineerde groepen.

907 Zie: Kamerstukken II 2020-2021, 26 643, nr. 765, actie 6.

Organisaties als de FNV, de Cliëntenraad, Artikel 1 en de Woonbond doen belangrijk werk voor specifieke doelgroepen in de samenleving. AI biedt deze groepen nieuwe mogelijkheden, maar kan hun positie en die van hun achterban ook onder druk zetten en zelfs schade toebrengen. Denk bij dit laatste aan een '*digital poorhouse*' voor de armen, een '*New Jim Code*' voor mensen van kleur en vrijheidsbeperking van minderheden door digitale 'openluchtgevangenissen'. Het is noodzakelijk dat juist ook de organisaties die dergelijke groepen vertegenwoordigen, in staat zijn om die effecten van AI te begrijpen en te adresseren. De specifieke kennis van deze partijen is bovendien onmisbaar bij de verdere inbedding van AI in de samenleving, maar ontbreekt momenteel in veel AI-discussies. Een belangrijke reden daarvoor is dat met name deze organisaties nog onvoldoende kennis over deze technologie bezitten.

De overheid is verantwoordelijk voor een sterke democratie en moet daarom zorg dragen voor een diversiteit van stemmen bij belangrijke vraagstukken. Bij de introductie van een nieuwe systeemtechnologie loopt het maatschappelijk middenveld doorgaans achter op het grote bedrijfsleven en de overheid, terwijl de stem ervan juist cruciaal is om misstanden rondom de technologie aan te kaarten en de technologie voor meer diverse doelen en belangen in te zetten. Het eerdergenoemde algoritmeregister kan deze achterstand verkleinen, door kennis over AI-gebruik openbaar toegankelijk te maken. Hiernaast is van belang dat de overheid belangenorganisaties zelf actief benadert en raadpleegt in het kader van haar AI-beleid.

De overheid kan echter ook bijdragen aan een prominentere rol van het maatschappelijk middenveld via de subsidies die zij verstrekt. Bovendien kan zij trainingen of samenwerkingsverbanden faciliteren. Maar ook de formele en institutionele mechanismen voor belangenbehartiging vragen in dit verband om aandacht. De WRR wijst hier op de noodzaak van een voldoende stevige betrokkenheid van ondernemingsraden (OR) en andere vormen van medezeggenschap. Zo bepaalt de Wet op de Ondernemingsraden dat werkgevers instemming van de OR nodig hebben voor het verwerken van persoonsgegevens van hun personeel. Toch staan de specifieke implicaties van AI, bijvoorbeeld bij de inzet van personeelsvolgsystemen, vaak nog onvoldoende op het vizier van een OR.

Concrete acties bij aanbeveling 5

- Betrek AI-geletterdheid in het subsidiebeleid en bij trainingen.
- Stimuleer samenwerking tussen belangengroepen en -organisaties in het digitale domein.
- Attendeer partijen in de samenleving op de diverse mogelijkheden, bijvoorbeeld door medezeggenschap, om betrokken te geraken bij besluiten over de inzet van AI.
- Betrek belangenorganisaties structureel en vroegtijdig bij de politieke besluitvorming over AI-beleid en -regels.

Het tweede punt bij de transitie van monoloog naar dialoog betreft de noodzakelijke aandacht voor een goede terugkoppeling tussen praktijk en tekentafel. Veel aandacht gaat uit naar de kwaliteit en betrouwbaarheid van data die bij AI-systemen worden gebruikt. Er is ook de nodige aandacht voor verschillende analysemethoden, de werking van de systemen en de transparantie daarvan. Anders gezegd, er is veel oog voor input en verwerking, maar minder voor output: doet het systeem wat het moet doen en zijn we daar tevreden mee?[908] Kennis over de uitkomsten van AI-systemen en de terugkoppeling daarvan naar de ontwikkelaars en andere betrokkenen klinken als een logische vereiste voor AI-systemen, maar toch gebeurt dit in de praktijk onvoldoende. Onze tweede aanbeveling bij de opgave van engagement is daarom:

Aanbeveling 6
Draag zorg voor een goede terugkoppeling tussen de ontwikkelaar van AI, de gebruiker ervan en de personen die er in de praktijk de consequenties van ondervinden.

Het relatieve gebrek aan aandacht voor terugkoppeling kent diverse oorzaken. Zo worden regelmatig '*real life*-experimenten' gehouden zonder dat de betrokkenen daar expliciet toestemming voor hebben gegeven. Na de experimentele fase worden deze systemen dan zonder verdere inhoudelijke evaluatie in werking gesteld. Relevant hierbij is dat AI veelal werkt met data over groepen personen, met als gevolg dat niet altijd sprake is van de verwerking van persoonsgegevens. Daarmee hebben belangrijke wettelijke waarborgen als het geven van toestemming voor datagebruik en compensatie bij misstanden

908 Vergelijk Kamerstukken II 2019-2020, 26643, nr. 641, in het bijzonder paragraaf 3.2 over kwaliteitswaarborgen.

sterk aan betekenis ingeboet.[909] Problematisch hierbij is ook dat dergelijke toepassingen tot groepsdiscriminatie kunnen leiden, terwijl er nog weinig mogelijkheden voorhanden zijn om als groep hiertegen in het geweer te komen. Ook kan het – vooral bij de overheid – zijn dat er akkoord wordt gegeven op de functionaliteiten en de werking van een systeem, maar dat op basis van het lerend vermogen van het systeem en de terugkoppeling erop aanpassingen doorgevoerd moeten worden die vereisen dat het hele systeem opnieuw beoordeeld wordt. Dergelijke lange en complexe processen bemoeilijken het leren van terugkoppeling.

Bij de overgang van het lab naar de samenleving veranderen bovendien de eisen ten aanzien van de terugkoppeling van AI-systemen. In het lab worden systemen vaak uitgebreid getest aan de hand van zorgvuldig samengestelde sets van testdata. Wanneer deze systemen in de samenleving gebruikt gaan worden, ontbreekt vaak de grondige controle en terugkoppeling van de resultaten die eigen is aan de gecontroleerde omgeving van onderzoekslaboratoria.

Weer een andere oorzaak van het ontbreken van terugkoppeling is dat de informatie daarvoor in sommige gevallen moeilijk te verkrijgen is. Een algoritme voor werving en selectie ontbreekt het aan terugkoppeling over mensen die onterecht afgewezen zijn, omdat er geen data zijn over hoe zij op het werk zouden functioneren. Algoritmen die schooladviezen geven, hebben voor de terugkoppeling data nodig die pas jaren later ontstaan. En zelfs dan is er veel onduidelijkheid: wat als iemand zich op latere leeftijd bijschoolt? Als iemand uiteindelijk een hoger diploma behaalt, was het algoritme dan verkeerd of ging de leerling pas later hard werken?

Ten slotte staan commerciële belangen van ontwikkelaars of contractuele afspraken tussen ontwikkelaars enerzijds en toepassende organisaties anderzijds een goede terugkoppeling in de weg. Om terugkoppeling te faciliteren en tegelijkertijd een bepaalde geheimhouding te garanderen wordt wel gesuggereerd dat het goed is een beperkt aantal personen binnen de toepassende organisatie aan te wijzen die kennis kunnen nemen van alle voor de terugkoppeling relevante factoren (type 'Commissie-Stiekem').

Een effectieve terugkoppeling is cruciaal voor het goed functioneren van AI-systemen en het beschermen van publieke waarden. De toeslagenaffaire is een tragisch voorbeeld van een gebrekkige terugkoppeling van en kritische reflectie op de uitkomsten van een systeem. Naarmate algoritmegebruik grotere implicaties krijgt voor de (rechts)positie van burgers, is het cruciaal dat signalen

909 Kosta 2020.

over de (uit)werking ervan actief georganiseerd worden. Die terugkoppeling dient een dubbele afstand te overbruggen. Enerzijds gaat het om terugkoppeling tussen ontwikkelaar van het AI-systeem en de gebruiker, bijvoorbeeld een huisarts, een politieagent of een leraar. Te vaak nog zitten er schotten tussen deze twee groepen. Daarnaast is er terugkoppeling nodig met de personen die door de systemen geraakt worden, bijvoorbeeld de patiënt, de aangehouden burger of de leerling. Zowel gebruiker als aan het AI-systeem onderworpen personen zijn in een positie om fouten te herkennen, expertise in te brengen en verbeteringen voor te stellen. In plaats van eenrichtingsverkeer is er dus dialoog nodig.

De overheid zal dus meer aandacht moeten besteden aan de wijze waarop en de mate waarin terugkoppeling georganiseerd is, vooral in de publieke sector, bij gemeenten, uitvoeringsinstanties en met name in domeinen waar beslissingen grote gevolgen voor burgers hebben. Naar het oordeel van de WRR dient een, nog te ontwikkelen, standaard voor terugkoppeling hier feitelijk verplicht te zijn.

Concrete acties bij aanbeveling 6

- Breng in verschillende domeinen in kaart wie ontwikkelaar, gebruiker en aan AI-systemen onderworpen burgers zijn en ontwikkel goede mechanismen voor terugkoppeling.
- Stel bij overheidstoepassingen terugkoppeling verplicht.
- Organiseer terugkoppeling in gebieden met gevoelige informatie op getrapte wijze.

9.5 Transitie 4: Van reactie naar regie

Bij de benadering van regulering is een transitie van reactie naar regie vereist. Met reactie doelen wij op een primair afwachtende houding op het terrein van wetgeving, waarbij de wetgever in actie komt om acute maar veelal specifieke vraagstukken te adresseren. Het risico daarbij is dat de wetgever zowel de betekenis van de bredere effecten van AI op de samenleving uit het oog verliest als de verschillende dossiers onvoldoende vanuit een bredere samenhang benadert. Zaken als betrouwbaarheid, uitlegbaarheid en transparantie zijn absoluut belangrijk, maar van doorslaggevend gewicht zijn eveneens de inrichtingsvraagstukken waar een samenleving met AI voor komt te staan. Met de transitie naar 'regie' bepleiten wij dat de wetgever op korte termijn een actievere rol pakt, ontwikkelingen veel meer vanuit een integraal perspectief adresseert en tijdig via regulering stuurt op de economische en maatschappelijke context waarbinnen AI tot wasdom komt. Naast het reguleren van de *werking* van de technologie zal de wetgever ook moeten inzetten op de regulering van de dynamiek en andere economische wetmatigheden waarmee AI gepaard gaat, zoals

de groeiende machtsconcentratie van een beperkt aantal, veelal private partijen en de gevolgen daarvan voor de plek die AI in de samenleving krijgt. In feite gaat het bij deze transitie om een regisserende rol van de overheid bij de inrichting van onze 'digitale leefomgeving'. Onze eerste aanbeveling voor de transitie van reactie naar regie is daarom:

> Aanbeveling 7
> Koppel de regulering van AI aan een discussie over de inrichting van de digitale leefomgeving en stel een brede wetgevingsagenda op.

De afgelopen jaren zijn diverse reguleringsprocessen in gang gezet, zo zagen we in hoofdstuk 7. Het gaat hierbij om zowel nationale als internationale initiatieven, van Europese wetgeving over AI en datagebruik en discussies over gezichtsherkenning en autonome wapens tot de regels en richtlijnen die de Nederlandse ministeries opstellen. Met de Europese concept-Verordening AI ligt er inmiddels een concreet voorstel voor een AI-regime gebaseerd op verschillende risicocategorieën, waar ook Nederland op termijn aan is gehouden. Deze reguleringsprocessen vinden overwegend plaats rondom acute en relatief helder afgebakende vraagstukken, zoals het gebruik van algoritmen voor fraudebestrijding, bias en discriminatie, transparantie en onbetrouwbare uitkomsten. Het debat over regulering richt zich hierbij overwegend op vragen over de relevantie van bestaande kaders dan wel de noodzaak van nieuwe regulerende en toezichthoudende instituten. Wat te weinig aan de orde komt, is de vraag welke publieke waarden we actief willen garanderen dan wel realiseren en welke stappen dat concreet vereist voor de inbedding van AI.

Naarmate AI meer ingebed raakt in de samenleving, zullen namelijk ook tweede- en derde-orde-vraagstukken optreden waarvoor nieuwe regels nodig zijn. Het zijn bij een systeemtechnologie niet alleen de concrete toepassingen die vragen oproepen, maar vooral de andere economische dynamieken en bredere effecten ervan in de samenleving. Ook elektriciteit en de komst van de auto noodzaakten de wetgever ertoe om de ontwikkelingen te beschouwen vanuit de bredere effecten daarvan op de samenleving, in dit geval de fysieke leefomgeving. Denk aan de aanleg van kabels boven of onder de grond en de aanleg van een wegenverkeersnet in afweging met de natuur. Bij de inbedding van AI spelen soortgelijke keuzes over de inrichting van de, digitale, leefomgeving. Deze leefomgeving omvat nu al zeer veel aspecten van de samenleving. De WRR wijst er daarom op dat de regulering van AI meer dient te omvatten dan alleen de kenmerken van de technologie en de toepassingen ervan (zijn deze betrouwbaar, veilig en transparant?). De overheid zal bij de regulering van AI hoe dan ook de bredere digitale leefomgeving moeten betrekken.

De ontwikkeling van eerdere systeemtechnologieën laat zien dat de rol van de overheid zal groeien naarmate AI meer ingebed raakt in de samenleving. Zij doet er daarom verstandig aan die grotere rol tijdig ter hand te nemen. De Europese concept-Verordening voor AI regelt de toelating van AI-toepassingen in de lidstaten, maar laat tegelijkertijd nog veel open omdat de insteek vooral is gericht op risicobeheersing. De WRR erkent dat het potentieel alomtegenwoordige karakter van AI het lastig maakt op voorhand te overzien welke kaders onder druk komen te staan of anderszins aanpassing behoeven. Dit zal in veel gevallen pas in de loop van de tijd duidelijk worden. Toch mag de wetgever niet afwachten, daarvoor zijn de (publieke) belangen die op het spel staan te groot. De wetgever zal greep op de ontwikkelingen moeten blijven houden, anders verliest hij het vermogen om die ontwikkelingen op tijd te kunnen bijsturen. Hiertoe zal de overheid niet alleen moeten investeren in onderzoek en signalering, zoals nu gebeurt door bijvoorbeeld toezichthouders, maar ook concrete stappen moeten durven zetten. Het nieuwe en daarmee deels onzekere karakter van AI moet daarbij niet worden overschat. Reeds bekende onduidelijkheden en spanningen met de bestaande wettelijke kaders kunnen eenvoudigweg worden verhelderd of uit de weg worden geruimd. Dit zal de verdere ontwikkeling en inbedding van AI in de samenleving alleen maar ten goede komen. Onduidelijkheid over de toepasbaarheid van de bestaande kaders en rechtsonzekerheid, waaronder ook wat betreft de bij AI te gebruiken data, staan een brede toepassing van AI immers in de weg. Keuzes hierover kunnen beter nu al gemaakt worden, omdat we als samenleving anders voor voldongen feiten kunnen komen te staan.

Het meest urgent is wat betreft de WRR dat de overheid het initiatief naar zich toe trekt bij de langetermijnontwikkeling van AI en richting geeft aan de brede maatschappelijke effecten ervan. Dan gaat het om kwesties als de doelen die we als samenleving willen nastreven en de vraag waar, waarvoor en onder welke condities we AI willen gebruiken. Hiertoe behoort ook het beperken en zelfs verbieden van het gebruik van AI op bepaalde terreinen, zoals de Europse Verordening voor AI voorstelt. Maatschappelijke kansen verdienen daarbij in het bijzonder de aandacht. Net als elektriciteit is AI behalve een economisch goed ook iets waarvan de baten ten goede komen aan grote groepen in de samenleving of zelfs de gehele bevolking. De komst van elektriciteit verlengde de dag, maakte woningen veiliger, steden schoner en veraangenaamde het leven in veel opzichten. Van AI mogen soortgelijke baten worden verwacht. Dan zal de overheid er wel zorg voor moeten dragen dat de technologie op plekken terecht komt waar ze een optimale bijdrage kan leveren, en wordt ingezet op een schaal en voor doelen die congruent zijn met die van de Nederlandse samenleving.

De discussie hierover vereist een afweging tussen verschillende publieke waarden en die kan en mag niet uitsluitend worden gemaakt door de technische experts en technologiebedrijven. Wellicht meer nog dan de vraag of de

bestaande kaders voldoende zijn of dat AI nieuwe regels vereist, is het dus van belang dat de overheid zich een beeld vormt van de inrichtingsvraagstukken rond de inbedding van deze systeemtechnologie en de rol die overheidsregulering daarin heeft te spelen.

De WRR bepleit dat een dergelijke, strategischer benadering van AI plaatsvindt op de wijze waarop de rijksoverheid dat bijvoorbeeld tot voor kort deed op het terrein van de ruimtelijke ordening, namelijk via diverse nota's. Behalve op de systematiek van de *Nota's ruimtelijke ordening* kan de overheid ook teruggrijpen op de *Nota Wetgeving voor de elektronische snelweg* uit 1998, waarin een strategische visie op het internet werd uitgewerkt middels diverse beleidsopgaven en beleidsdoelen, de daarbij horende sturingsfilosofie, het instrumentarium en de uitvoering (zie ook tekstbox 9.6).

Tekstbox 9.6 – De Nota Wetgeving voor de elektronische snelweg (1998)

Met de *Nota Wetgeving voor de elektronische snelweg* presenteerde de overheid in 1998 haar zienswijze op de regulering van het internet (de elektronische snelweg). Ten grondslag aan de nota lag een uitgebreide studie van de belangrijkste gevolgen van de elektronische snelweg voor de Nederlandse wetgeving. Deze studie omvatte behalve technische, bestuurskundige en juridische verkenningen en een internationale rechtsvergelijking van het internet ook een bespreking van strategische thema's als internationalisering en rechtsmacht, betrouwbaarheid, markten en rechtshandhaving.

Een toetsingskader voor de wetgever, om actoren in het wetgevingsproces houvast te geven bij vragen van wetgeving rond de elektronische snelweg, en een reeks voorstellen voor het opstellen, aanpassen en intrekken van wetgeving en voor de inbreng vanuit Nederland in internationale overlegfora vormden de kern van de nota. Voor de implementatie van deze voorstellen bevatte de nota een actieplan, met daarin een prioriteitstelling.

Concrete acties bij aanbeveling 7

- Accepteer dat wetgeving gericht op de inbedding van AI een langdurig en deels onzeker proces zal zijn. Pas het wetgevingsinstrumentarium daarop aan, maar wacht niet af met ingrijpen.
- Stel een brede en integrale wetgevingsagenda op voor AI en de inrichting van de digitale leefomgeving, met daarin gespecificeerd beleidsdoelen, de daarbij horende sturingsfilosofie, het instrumentarium en de uitvoering.
- Neem in deze agenda een lijst op van wettelijke bepalingen waarvan op korte termijn de betekenis voor AI moet worden geëxpliciteerd. Te denken valt aan bepalingen over geautomatiseerde besluitvorming, aansprakelijkheid, archivering en de juridische status van autonoom handelende systemen.
- Versterk de signalerende rol van toezichthoudende instanties en zorg voor een terugkoppeling naar beleid en wetgeving. Verwerk deze signalen, samen met die van andere partijen, eventueel in een aparte monitor.

De tweede aanbeveling voor de transitie van reactie naar regie betreft de concrete focus van de overheid bij de regulering van AI als systeemvraagstuk:

Aanbeveling 8
Stuur via wetgeving actief op ontwikkelingen rondom surveillance en dataverzameling, de scheve verhouding tussen publiek en privaat in het digitale domein en machtsconcentratie.

Door de regulering van AI te bezien als systeem- en daarmee inrichtingsvraagstuk komt in beeld welke ontwikkelingen de inrichting van de 'digitale leefomgeving' bepalen. Wanneer de overheid niet nu al actief stuurt op hoe en door wie AI in de samenleving gebruikt gaat worden, is de kans aanwezig dat zij op termijn daarop haar greep verliest. Op ten minste drie terreinen zijn hiertoe stappen vereist.

Allereerst is het nodig om de afhankelijkheid van de publieke sector van private bedrijven te verkleinen. Terwijl het gebruik van AI in de private sector toeneemt, stelt de overheid zich vanwege onwennigheid en groeiende zorgen terughoudend op. We noemen een paar voorbeelden van deze kloof. Op het gebied van veiligheid dient de politie zich aan strenge regels te houden. Maar wat als er private diensten of applicaties opkomen waarmee burgers zelf gezichtsherkenningssoftware gaan gebruiken voor opsporing? Of neem de

openbare ruimte. Voor het handelen van burgers en bedrijven aldaar is de overheid primair verantwoordelijk. Tegelijkertijd is er inmiddels een wildgroei van partijen die binnen en over de publieke ruimte informatie verzamelen met camera's, drones en sensoren – en daarmee potentieel over meer informatie beschikken dan de overheid zelf. Als gevolg van deze kloof kan de overheid grip verliezen op gebieden waar zij verantwoordelijk voor is, temeer omdat met het gebruik van AI door derde partijen ook de voor het beleid noodzakelijke kennis wegsijpelt. Daar komt nog bij dat heikele vragen zo uitbesteed kunnen worden en dubieuze praktijken aan het zicht onttrokken worden.

Ten tweede zal de groei van massasurveillance en daarmee het veelal ongericht verzamelen en (her)gebruiken van data een halt moeten worden toegeroepen. Natuurlijk kan ook hier – zoals momenteel gebeurt – van geval tot geval gekeken worden hoe de maatschappelijke kosten van surveillance en datagebruik zich verhouden tot de baten ervan en zijn per applicatie (variërend van gezichtsherkenning en onlinegedragsbeïnvloeding tot intelligente toepassingen in woningen[910]) allerlei waarborgen te stellen. Maar de ontwikkeling kent ook een structurelere kant. Het volgen van mensen, in al hun gedragingen en zelfs emoties of unieke DNA-kenmerken, is inmiddels een belangrijk bestanddeel van het verdienmodel van talloze bedrijven, waaronder onlineplatforms.[911] Vormen van surveillance liggen meer en meer ten grondslag aan een groot deel van de interneteconomie. AI is een volgende fase in deze ontwikkeling, omdat bedrijven hiermee nog beter in staat zijn individuen te volgen, te profileren en op hun voorkeuren in te spelen. Relevant zijn hier ook de sterke toename en verspreiding in het dagelijkse leven van het aantal digitale apparaten die het volgen van mensen verder faciliteren. Precies om die reden begeven de grote technologiebedrijven zich op de markt voor slimme consumentenelektronica of gaan zij allianties aan met bedrijven in de maakindustrie. Surveillance – ook zoals de overheid deze zelf heeft ingezet – is van grote invloed op het gebruik en de perceptie van AI, en roept vragen op over op hoe bedrijven data gebruiken en overheden zich tot burgers verhouden. Als publieke waarden als privacy, individuele autonomie, veiligheid en democratische controle niet eerst beter worden gehandhaafd en geborgd, kan AI nooit een legitieme plek in de samenleving verwerven.

Problematisch voor de verdere ontwikkeling van AI is tot slot de zeer grote machtsconcentratie bij een beperkt aantal technologiebedrijven. Het gaat hier vooral over Amerikaanse bedrijven als Google, Facebook, Amazon, Microsoft en Apple, tevens de grootste spelers op het terrein van AI. Als gevolg van de

910 De Conca 2021.
911 Rathenau Instituut 2021.

coronacrisis en de groei van thuiswerken en videoconferencing is hun macht verder vergroot. Een studie van de Werkgroep Publieke Waarden in het onderwijs van de VSNU toont bijvoorbeeld de zeer grote concentratie van slechts enkele leveranciers in verschillende onderdelen van het hoger onderwijs in Nederland.[912] Wereldwijd groeit de onvrede met de macht van de bedrijven uit Silicon Valley en komen overheden in actie. Onder meer de Europese Commissie, het Amerikaanse ministerie van Justitie en de Britse regering hebben deze bedrijven betiteld als een bedreiging voor innovatie, mededinging en privacy. Daar komt volgens hen bij dat de wijze waarop deze bedrijven informatie filteren en verspreiden, steeds meer wordt gezien als een ernstige politieke dreiging, niet alleen voor kwetsbare democratische regimes, maar ook voor oude en gevestigde democratieën als de Nederlandse.

De grote technologiebedrijven zijn in staat om langdurig en verstrekkend richting te geven aan de ontwikkeling en het gebruik van AI. Hun enorme omvang en middelen stellen hen daartoe in staat. Door netwerkeffecten kunnen zij bovendien een centrale positie in andere sectoren verwerven. Niet democratische waarden maar commerciële belangen zijn daarbij leidend. Een dergelijke positie en macht van grote technologiebedrijven is in het bijzonder problematisch wanneer de diensten die zij leveren tot de maatschappelijke infrastructuur gaan behoren. Hoe de macht van de grote technologiebedrijven beperkt zal worden, is vooralsnog onduidelijk, maar de geschiedenis van systeemtechnologieën leert dat monopolies ofwel opengebroken ofwel gedwongen worden hun infrastructuren open te stellen voor anderen. Hiervoor circuleren inmiddels verschillende voorstellen.[913] De meest concrete daarvan zijn afkomstig van de Europese Commissie, die met de Digital Markets Act en de Digital Services Act de contouren schetst van een samenhangend Europees internetrecht.[914] De WRR ondersteunt dat Nederland aan deze voorstellen actief bijdraagt, en daarvoor de noodzakelijke input levert. Nederland kan op dit terrein echter ook zelf stappen ondernemen, door de eigen wetgeving en het toezicht op het terrein van mededinging aan te passen met het oog op datamacht. Ook het bij aanbeveling 3 genoemde actiepunt om het aanbestedingsbeleid beter te benutten, kan een middel zijn om diversiteit van aanbieders van producten en diensten te stimuleren.

Naast deze voorstellen gericht op machtsbeperking en een goed werkende markt zijn er echter ook initiatieven om minder afhankelijk te worden van private toeleveranciers en met publieke middelen alternatieven te bouwen. Een voorbeeld

912 VSNU, 16 april 2021.
913 CPB 2021, p, 16.
914 Chavannes et al. 2021.

zijn de in hoofdstuk 4 en 8 genoemde initiatieven om eigen clouddiensten en Europese AI-centra te ontwikkelen – onder andere in Nederland – en het project AI4EU. Op kleine schaal zijn er ook verdergaande initiatieven, zoals de instelling van digitale nutsvoorzieningen op onder meer het terrein van elektronische identificatie. Een dergelijke nutsvoorziening zou ook voor AI verkend kunnen worden, bijvoorbeeld als onderdeel van de eerdergenoemde AI-identiteit en de ondersteunende technische infrastructuur die daarbij hoort. Het grote belang van dergelijke initiatieven ligt erin dat ze op basis van publieke waarden kunnen worden ontwikkeld. Dit is bij uitstek relevant voor publieke sectoren als de zorg en het onderwijs, die immers op publieke diensten moeten kunnen draaien.

Waar het de WRR bij de transitie van reactie naar regie vooral om gaat, is dus dat de overheid zich rekenschap geeft van het feit dat regulering van AI alleen niet volstaat. Zij zal ook op veel andere terreinen maatregelen moeten treffen om ervoor te zorgen dat de omgang met AI in lijn blijft met de publieke waarden of die versterkt. Als dat besef ontbreekt en de overheid niet tijdig anticipeert op deze bredere opgave, bestaat de kans dat andere belangen en partijen bij de inbedding van AI de boventoon gaan voeren. Vanwege het in hoofdstuk 7 besproken moment van *closure* is bijsturing dan nauwelijks meer realistisch.

Concrete acties bij aanbeveling 8

- Garandeer en borg de zeggenschap over kritieke digitale voorzieningen en bouw deze waar nodig zelf op, onder meer op de domeinen van de Nederlandse AI-identiteit en publieke sectoren als zorg of onderwijs.
- Bezie wetgevingsbeleid inzake surveillance mede vanuit het besef dat AI een volgende fase in de ontwikkeling van surveillance is.
- Maak veel meer gebruik van het aanbestedingsinstrumentarium om publieke waarden te borgen; voorkom dat het uitsluitend grote technologiebedrijven in de kaart speelt.
- Draag actief bij aan Europese wetgeving en daaraan verbonden initiatieven om AI en de bredere digitale leefomgeving te reguleren.
- Maak haast met de aanpassing van het mededingingsrecht met het oog op de data-economie en AI-bedrijven.

9.6 Transitie 5: Van natie naar netwerk

Positionering ten slotte vraagt om een transitie van natie naar netwerk. Dat betekent dat we het verdienvermogen met AI niet louter als een *zero sum*-competitie met andere landen moeten beschouwen, maar ook moeten werken aan betere verbindingen met andere landen. In het bijzonder geldt dat voor de landen binnen de EU. Daarnaast betekent de transitie dat veiligheid niet alleen een

zaak is van externe dreigingen aan de landsgrenzen, maar ook samenhangt met technologieën die burgers in hun dagelijks leven gebruiken. Om zicht te krijgen op veiligheidsdreigingen moeten we ook hier de aandacht verleggen naar het internationale netwerk waarmee we verbonden zijn. De WRR bepleit dat we het idee verlaten dat Nederland in een competitie verwikkeld is met andere landen om welvaart en macht (natie) en ons richten op de verbindingen die wij met het buitenland hebben (netwerk). Ten aanzien van de economische kant van die opgave is onze aanbeveling:

Aanbeveling 9
Versterk het Nederlandse verdienvermogen met een 'AI-diplomatie' die gericht is op internationale samenwerkingsverbanden, in het bijzonder binnen de EU.

Overheden en bedrijven doen wereldwijd veel om hun verdienvermogen met AI te versterken. Op tal van gebieden is er sterke concurrentie tussen landen. Denk aan de omvang van publieke en private investeringen, maar ook aan het ontwikkelen en behouden van talent. Het is belangrijk dat Nederland daar oog voor heeft, omdat veel van onze buurlanden daar inmiddels flink op inzetten. In het SAPAI komt dit beleid aan bod onder de noemer van het 'benutten van de kansen'. Er is onder experts en belanghebbenden in dit verband discussie over de vraag of Nederland wel voldoende doet om in 'de race' mee te blijven doen. De Nederlandse AI Coalitie is mede gericht op het bij elkaar brengen van private en publieke partijen om gezamenlijk vast te stellen wat nodig is om ons verdienvermogen te versterken.

De WRR bepleit een ander perspectief op de rol en positie van Nederland. Er dient meer aandacht te zijn voor het versterken van het verdienvermogen door internationale samenwerking, niet primair door concurrentie maar door een aanpak van integrale 'AI-diplomatie'.

Aan wat voor internationale samenwerking moeten we dan denken? Een eerste domein waar dat op kan plaatsvinden, is fundamenteel onderzoek. Het Europese verband CLAIRE heeft Den Haag als zetel gekozen.[915] Het versterken van dergelijke verbanden in Nederland kan diverse spin-offs voor het Nederlandse bedrijfsleven creëren. Een analogie daarvoor is het Zwitserse CERN, waarbij gepoold onderzoek Europa vooraanstaand heeft gemaakt op het gebied van de deeltjesfysica. Het is verstandig om te leren van de voorwaarden

915 Zie hiervoor tekstbox 8.1 in hoofdstuk 8.

waaronder dergelijke samenwerkingen op het gebied van onderzoek succesvol zijn.[916]

Samenwerking kan ook plaatsvinden op de ontwikkeling van concrete AI-toepassingen. Frankrijk en Duitsland hebben bijvoorbeeld het initiatief genomen voor een Europese data- en clouddienst onder de naam Gaia-X.[917] Nederland sloot zich na enige tijd aan bij dit initiatief, en inmiddels is er ook een Nederlandse hub om op Europees niveau de nationale belangen naar voren te kunnen brengen. Critici wijzen op de onhaalbaarheid van dergelijke projecten, maar er is een historie van succesvolle Europese samenwerking op het gebied van technologie met bijvoorbeeld Galileo, het Europese alternatief voor GPS en het luchtvaarbedrijf Airbus. Ook hierbij is het verstandig om te leren van succes en falen in het verleden bij dergelijke samenwerkingsverbanden.[918] Dit samenwerkingsverband heeft zeker het potentieel om de Europese positie te versterken. Wanneer Nederland niet participeert, zullen de Nederlandse belangen sowieso niet meegenomen worden.

Samenwerking voor verdienvermogen kan ook de vorm aannemen van meer coördinatie in het veld rondom bestaande bedrijven. De groeiende verwevenheid tussen economie en geopolitiek heeft ertoe geleid dat allerlei digitale technologieën onderwerp van handelsconflicten werden. Nederlandse bedrijven als ASML en NXP, belangrijk voor de hardware van AI-toepassingen, zijn al onderhevig aan de grillen van de Amerikaans-Chinese handelsrelatie. Vergelijkbare situaties kunnen zich in de toekomst voordoen bij Nederlandse technologiebedrijven als Philips, KPN, TomTom of Adyen. Nederland doet er goed aan bij te dragen aan een Europese positionering om de positie van individuele landen en bedrijven te versterken in de competitie tussen grootmachten. Specifiek gaat het om beleid in het geval van (vijandige) overnames, boetes of verboden die handelspartners kunnen opleggen.

Weer een andere dimensie waarlangs samenwerking het verdienvermogen van Nederland kan versterken, is wet- en regelgeving. De EU was en is actief op het gebied van persoonsgegevens (de AVG) en de genoemde concept-Verordening voor AI van april 2021.[919] Hiernaast is het proces van standaardisatie cruciaal. Dit technische veld krijgt binnen het debat over AI relatief weinig aandacht, maar is immens invloedrijk en heeft grote effecten op de competitiviteit van

916 Zie bijvoorbeeld: Smith 1999.
917 Zie hiervoor tekstbox 8.2 in hoofdstuk 8.
918 Domini en Chicot 2018.
919 Het onderzoek van Anu Bradford (2020) naar 'het Brussels Effect' laat zien hoe de wetgeving van de EU op verschillende terreinen mondiaal de toon zet. Als zogenoemde *'regulatory power'* kan de EU de richting van de markt beïnvloeden.

landen.[920] Bovendien is standaardisatie, zoals we in hoofdstuk 8 lieten zien, in toenemende mate onderhevig aan 'geopolitisering'. Met name China probeert in internationale fora de eigen standaarden voor AI tot norm te verheffen. De EU, en Nederland daarbinnen, dient hierop zeer alert te zijn en de samenwerking te zoeken met landen die dezelfde waarden onderschrijven.

Hoewel de EU op de meeste gebieden van samenwerking het aangewezen forum is, kan op specifieke dossiers ook de samenwerking worden gezocht met individuele gelijkgezinde en pionierende landen zoals Canada, Frankrijk, Zuid-Korea of Singapore. Maar bij vraagstukken rondom digitalisering is het sowieso belangrijk oog te houden voor mogelijke brede coalities van landen.[921]

Concrete acties bij aanbeveling 9

- Inventariseer verschillende domeinen en fora voor samenwerking op het gebied van AI.
- Kijk per domein waar de kansen liggen om Nederland te positioneren.
  Doe dat in samenhang met de eerder besproken Nederlandse 'AI-identiteit'.
- Betrek nationale en internationale actoren, zoals standaardisatie-instituut NEN en prominente wetenschappers, bij de formulering van beleid.
- Formuleer doelen per domein, maar ook synergieën over domeinen heen, tussen bijvoorbeeld fundamenteel onderzoek en Europese projecten voor AI-toepassingen.
- Wees alert op de AI-diplomatie die schuilgaat achter reguleringsvoorstellen van andere landen en die Nederlandse belangen kan schaden.

Kortom, op de gebieden van fundamenteel onderzoek, het opzetten van nieuwe diensten, coördinatie rondom bedrijven en wet- en regelgeving kan internationale samenwerking bijdragen aan het verdienvermogen van Nederland. Om op deze domeinen weloverwogen keuzes, mede gericht op langetermijnbelangen, te kunnen maken, adviseert de WRR daarom een integrale AI-diplomatie te ontwikkelen.

920 Veale en Zuiderveen Borgesius 2021.
921 Zie ook: WRR 2015.

De transitie van natie naar netwerk heeft behalve een economische ook een veiligheidsdimensie. Onze aanbeveling daarvoor luidt:

> Aanbeveling 10
> Weet je als land ook in het AI-tijdperk te verdedigen; versterk daarom de Nederlandse capaciteiten tegen de groeiende 'informatieoorlog' en de export van digitale dictatuur.

Bij de vraag naar de invloed van AI op de veiligheid gaat de discussie al gauw over autonome wapens. Die kunnen gevaarlijke gevolgen hebben en de huidige bemoeienissen hiermee zijn dan ook toe te juichen. AI beïnvloedt het militaire domein echter ook op andere manieren, namelijk via betere besluitvorming of het analyseren van meer data. De aandacht hiervoor neemt toe, bijvoorbeeld in NAVO-verband. De WRR benadrukt het belang van een bredere blik. Veiligheid betreft namelijk niet alleen AI in het militaire domein, maar ook in het civiele domein.

De vergaande digitalisering van samenleving en economie maakt ons land namelijk ook kwetsbaar voor aanvallen langs deze weg. Socialemediaplatformen, sensoren in de infrastructuur, besturingssystemen en communicatiesystemen en andere 'vernetwerkte' domeinen zijn allemaal potentiële kwetsbaarheden. Cybersecurity is inmiddels een sterk groeiend beleidsterrein. In een recent rapport bepleitte de WRR meer aandacht voor de voorbereiding op het fenomeen van digitale ontwrichting. Naast de infrastructuur en de netwerken is er meer aandacht nodig voor de informatie die daardoor stroomt.[922] De beïnvloeding en manipulatie van informatie maken dat er een wat wel 'informatieoorlog' wordt genoemd, is. Die geschiedt deels handmatig, maar ook steeds vaker middels algoritmen.

De WRR wijst op de noodzaak om dit risico integraal te benaderen. Lange tijd was de veronderstelling dat digitale technologieën een inherente democratiserende werking hebben. Alhoewel zij daar zeker aan kunnen bijdragen, zijn verschillende autoritaire regimes zeer capabel gebleken om die technologieën ook voor hun doelen in te zetten. Digitalisering, en AI in het bijzonder, kan autoritaire regimes zelfs versterken door surveillancetechnieken bijvoorbeeld grootschalig, gecentraliseerd en goedkoop beschikbaar te stellen. Daar komt bij dat landen als China en Rusland dergelijke technologieën in toenemende mate exporteren. Dat heeft effecten voor derde landen, die hierdoor in een meer

922 WRR 2019.

autoritaire richting kunnen gaan bewegen. Maar de risico's kunnen uiteindelijk ook ons land treffen. Door digitalisering en AI te beschouwen als middelen voor de nationale veiligheid, hebben dergelijke autoritaire landen grote capaciteiten opgebouwd. Daarover is wat betreft de WRR in ons land meer bewustwording nodig. De discussie dient bovendien breder te zijn dan slechts over de risico's rondom de uitrol van 5G en de mogelijke rol van bedrijven als Huawei daarbij. We noemen in dit verband de import uit het buitenland van bijvoorbeeld camera's met gezichtsherkenning, *smart city*-technologie die de openbare ruimte monitort en nieuwe telecominfrastructuur en software voor publieke diensten. Ook de export van technologie van Nederlandse bedrijven voor autoritaire doeleinden in het buitenland valt hieronder. Ten slotte betreft deze bedreiging campagnes om nepnieuws, *deepfakes* en samenzweringstheorieën in ons land te verspreiden (zie tekstbox 9.7).

Tekstbox 9.7 – AI als instrument van informatieoorlog

Microtargeting, *sentiment analysis* en *natural language processing* zijn technieken die steeds meer ingezet worden, maar die de veiligheid van burgers in gevaar brengen. Een opkomende ontwikkeling zijn de zogenoemde *deepfakes*, video- en audiobeelden die steeds moeilijker van echt te onderscheiden zijn. Dat brengt risico's mee voor individuele burgers en de samenleving als geheel doordat ze wantrouwen, onzekerheid en chaos creëren. Dergelijke technologieën kunnen uiteindelijk ook een risico vormen voor de democratie.

In EU-verband zijn er initiatieven op dit gebied vanuit de groeiende zorg om 'digitale soevereiniteit'. In ons land vroeg de Cyber Security Raad (CSR) begin 2021 expliciet om een veel actievere opstelling van de Nederlandse overheid in deze om de controle op democratie, rechtsstaat en economisch innovatiesysteem te kunnen behouden.[923] In navolging van de CSR bepleit de WRR dat Nederland zich in dit veld inzet voor een gezamenlijke Europese strategie. Het initiatief van ons land om samen met Frankrijk en Duitsland op EU-niveau te pleiten voor een EU-toezichthouder die alle fusies en overnames door grote digitale platforms met een poortwachtersfunctie kan beoordelen, vormt hierin een goede eerste stap.[924]

Het is echter ook zaak dat Nederland zelf beter in kaart brengt op welke manier buitenlandse mogendheden informatie inzetten en hoe dat ons democratisch

923 Cyber Security Raad 2021.
924 Rijksoverheid, 27 mei 2021.

bestel onder druk kan zetten. Vervolgens dienen wij onze capaciteiten, waaronder AI-capaciteiten, te versterken om dit tegen te gaan. Het is niet op voorhand evident met welke middelen de zogenoemde informatieoorlog gewonnen zou kunnen worden. Duidelijk is wel dat er geen tijd te verliezen is wat betreft het opbouwen van expertise en eventuele beleidskeuzes. Een eerste stap die Nederland al op korte termijn kan zetten is door aan dit type bedreiging meer aandacht te besteden in het jaarlijkse cybersecuritybeeld.

Concrete acties bij aanbeveling 10

- Inventariseer hoe verschillende vormen van AI als *microtargeting* en *deepfakes* onderdeel zijn van een mondiale informatieoorlog.
- Begrens technologieën van digitale dictatuur die naar Nederland worden geëxporteerd en technologieën van Nederlandse bedrijven die elders voor dictatoriale doeleinden worden ingezet.
- Bouw verder aan een Nederlandse positie op het gebied van digitale soevereiniteit in EU-verband.
- Neem veiligheidsrisico's rondom informatie systematisch op in het jaarlijkse cybersecuritybeeld.

9.7 Van instrumentarium naar beleidsinfrastructuur

Bovenstaande aanbevelingen betreffen het werk dat verricht moet worden bij het inbedden van AI in de samenleving. Onze slotaanbeveling betreft de ondersteuning van dat werk en richt zich op de institutionele kant van het overheidsbeleid inzake AI. Hiermee geven we ook antwoord op het tweede deel van de adviesaanvraag. De regering verzocht de WRR namelijk allereerst om de impact van AI op publieke waarden te onderzoeken. Dat hebben wij gedaan door AI als systeemtechnologie te begrijpen en vandaaruit vijf opgaven van maatschappelijke inbedding te onderscheiden en deze in verband te brengen met publieke waarden. Daarnaast stelde de regering de vraag of aanvullende instrumenten nodig zijn "om de kansen van AI te faciliteren en de uitdagingen te beantwoorden". In het verlengde van onze analyse van AI als systeemtechnologie bepleiten we hiervoor een laatste noodzakelijke transitie, namelijk van de focus op het beleidsinstrumentarium naar de opbouw van een beleidsinfrastructuur.

De geschiedenis van systeemtechnologieën leert, zoals gezegd, dat de rol van de overheid op allerlei verschillende manieren gaandeweg toeneemt. De spoorwegen werden in het Verenigd Koninkrijk en de VS ontwikkeld door private partijen, maar mettertijd kreeg de overheid een actievere rol via wetgeving en in veel Europese landen ook als aanbieder van openbaar vervoer. Hetzelfde gebeurde bij elektriciteit, waarbij de overheid zorgdroeg voor de aanleg van netwerken. De aard van de overheidsrol varieert dus, maar duidelijk is dat de

omvang ervan groeit. En met die grotere rol ontstond telkens ook een beleidsinfrastructuur om deze nieuwe taken te coördineren en de bijbehorende verantwoordelijkheden in te vullen. Zo kreeg in Nederland Rijkswaterstaat het beheer van de autowegen toegewezen en kwamen er verscheidene nieuwe publieke instanties rondom de auto, zoals de Dienst Wegverkeer voor de registratie van voertuigen en rijbewijzen, de Inspectie Leefomgeving en Transport voor de veiligheid van taxi's, busvervoer en wegtransport, en het Centraal Bureau Rijvaardigheidsbewijzen voor het afnemen van rijexamens.

We verwachten eenzelfde patroon bij AI. Dat betekent dat de overheid bij het inbedden van AI in de samenleving meer zal moeten doen dan het ontwikkelen van nieuwe instrumenten. De komende jaren zal zij eveneens moeten bouwen aan een beleidsinfrastructuur. Indachtig deze transitie van de ontwikkeling van het instrumentarium naar ook een ondersteunende beleidsinfrastructuur, luidt onze slotaanbeveling:

Slotaanbeveling
Bouw een beleidsinfrastructuur voor AI op, te beginnen met een AI-coördinatiecentrum voorzien van politieke verankering middels een ministeriële onderraad.

De noodzaak van een beleidsinfrastructuur wordt steeds duidelijker. Evenals eerdere systeemtechnologieën zal AI een variëteit aan zowel sectorspecifieke als generieke publieke waarden raken. Met de tijd zullen de risico's maar ook de kansen voor die waarden scherper in zicht komen. Ook zal steeds vaker debat nodig zijn over de doelen die we als samenleving willen nastreven en de vraag waar, waarvoor en onder welke condities we AI willen gebruiken. Bovendien vraagt AI om internationale samenwerking tussen landen, in het bijzonder binnen de EU. Aldus zal de overheid steeds meer betrokken raken. In aanvulling hierop stellen we vast dat het strategische belang van AI steeds meer wordt onderkend, wat eveneens om een actieve overheidsrol vraagt. Met deze ontwikkelingen wordt ook de noodzaak van brede en algemeen beschikbare middelen duidelijk om dit proces van beleid en wetgeving te ondersteunen.

De discussie rondom een beleidsinfrastructuur voor AI wordt in feite reeds impliciet gevoerd. Zo zijn er voorstellen voor een ministerie voor Digitalisering[925], waar AI onderdeel van zou zijn. Ook is er de roep om een toezichthouder voor algoritmen op te richten. Diverse landen zijn het stadium

925 ROB 2021.

van gedachtenvorming inmiddels voorbij zijn en nemen concrete initiatieven om AI institutioneel in te bedden (tekstbox 9.8). De WRR acht het wenselijk dat ook ons land hiervoor concrete stappen zet.

Tekstbox 9.8 – Landen maken inmiddels keuzes voor institutionele inbedding van AI

Diverse landen hebben hun nationale AI-strategieën laten ontwikkelen door commissies die bestaan uit experts uit de academische wereld, het bedrijfsleven en de overheid (bijvoorbeeld België, Groot-Brittannië, Frankrijk en Duitsland). Een argument voor deze brede samenstelling is dat AI in de toekomst alle sectoren van de maatschappij zal beïnvloeden en hierdoor ook alle ministeries aangaat.

Waren deze commissies in het begin veelal tijdelijk van aard en extern aan de overheid, intussen zijn er permanente organisaties, in de vorm van adviesraden (bijvoorbeeld Oostenrijk en Singapore), een taskforce binnen de overheid (bijvoorbeeld Kenia en India) of een initiatief belast met AI, zoals het *National Robotics Initiative* in de VS, waaraan een reeks van overheidsorganisaties bijdraagt. De Verenigde Arabische Emiraten heeft als enige land een ministerie voor AI, om wereldwijd voorop te kunnen lopen met AI binnen sectoren als transport, gezondheidszorg, hernieuwbare energie en verkeer en vóór 2117 huizen te bouwen op Mars.

Het Verenigd Koninkrijk heeft voor de implementatie van de eigen AI-missie en de bijbehorende Data Grand Challenge een 'Office for Artificial Intelligence' in het leven geroepen. Het bureau maakt onderdeel uit van de ministeries van Digital, Culture, Media & Sport en het ministerie van Business, Energy & Industrial Strategy. Belangrijke wapenfeiten van het bureau zijn onder meer de Richtlijnen voor AI-inkoop en de Richtlijnen voor AI-gebruik in de publieke sector. The Office for Artificial Intelligence en de bredere overheid worden voor de ontwikkeling van AI-beleid bijgestaan door een onafhankelijke AI-Raad (Council for AI), waarin AI-deskundigen en leden afkomstig uit de industrie, de publieke sector en de academische wereld zitting hebben. De raad werkt tevens aan de publieke perceptie van AI.

Hoewel de overheden van diverse landen stappen zetten op het terrein van een beleidsinfrastructuur voor AI, ontbreekt daarvoor een blauwdruk. Er zijn verschillen die te maken hebben met de missie van deze instanties, hun samenstelling en bevoegdheid en vooral ook de inbedding ervan in de overheid.

Dat sommige landen een agentschap (Denemarken), minister (Noorwegen, Zweden, Duitsland, Italië) of staatssecretaris (Frankrijk, België) voor digitalisering of de digitale overheid hebben, speelt eveneens een rol. Helder is in ieder geval dat ons land zich wat betreft een Nederlandse beleidsinfrastructuur voor AI volop kan laten inspireren door keuzes die andere landen al hebben gemaakt (zie ook tekstbox 9.9).

Maar er speelt nog een andere ontwikkeling die de noodzaak van een beleidsinfrastructuur agendeert. De Europese concept-Verordening voor AI eist van de lidstaten dat zij één of meer nationale bevoegde autoriteiten aanwijzen om toezicht te houden op de toepassing en uitvoering van AI en één nationale toezichthoudende autoriteit aanwijzen als officieel contactpunt voor het publiek en andere actoren. Deze autoriteit vertegenwoordigt ook de betreffende lidstaat in de European Artificial Intelligence Board, het samenwerkingsmechanisme dat de verordening implementeert.

Kortom, ons land zal in ieder geval op dit punt een beleidsinfrastructuur vorm moeten geven. Wat betreft een aanvullende stap richting een beleidsinfrastructuur voor AI, acht de WRR het in deze fase voorbarig om een apart ministerie of een specifieke toezichthouder voor AI te bepleiten. Beide voorstellen kunnen in een latere fase zinvol blijken, maar de meerwaarde ervan is op dit moment onvoldoende helder, temeer daar overlap met bestaande partijen reëel is. Het kost daarnaast veel tijd, middelen en energie om dergelijke zware gremia op te zetten. Centralisatie kan bovendien een onrealistische verantwoordelijkheid creëren. Een algoritme-autoriteit bijvoorbeeld laten beoordelen wat wel en niet toelaatbaar is, vraagt om domeinkennis – van regels, praktijken, normen – op talloze gebieden, variërend van zorg tot mobiliteit en defensie. Gegeven ook de vroege fase van AI als systeemtechnologie, is nog onduidelijk welke vraagstukken een algemenere, overkoepelende benadering vereisen vanuit de overheid.

Tekstbox 9.9 – Aanzetten voor een Nederlandse AI-beleidsinfrastructuur

Ook ons land kent gremia die als onderdeel van een AI-beleidsinfrastructuur gezien kunnen worden. Zo hebben verschillende departementen inmiddels afdelingen, directies of aparte organisatieonderdelen die zich met digitalisering bezighouden, zoals de directie Digitale overheid en de directie Digitale economie.

Daarnaast zijn het ministerie van Economische Zaken en Klimaat, het ministerie van Binnenlandse Zaken en Koninkrijksrelaties en het ministerie van Justitie en Veiligheid recent een samenwerking aangegaan op

het terrein van digitalisering. Deze drie ministeries waren verantwoordelijk voor de eerste Nederlandse digitaliseringsstrategie uit 2018 en de actualisering daarvan in 2020 en 2021.

Hiernaast is er een interdepartementale werkgroep, die gewerkt heeft aan de visie van het kabinet op de impact die digitalisering heeft op de publieke waarden en mensenrechten en aan het SAPAI, en een interdepartementale werkgroep van rijksinspecties en markttoezichthouders op AI.

Ook kent Nederland een recent herzien stelsel van CIO-officers, dat onder meer is gericht op de rijksbrede 'digitale transformatie en technologisch gedreven innovatie' en het generieke actieplan informatiehuishouding met een regeringscommissaris daarvoor. Ten slotte is er sinds 2021 een permanente Commissie Digitale Zaken in de Tweede Kamer.

Dit betekent echter niet dat de WRR de huidige status quo van het overheidsbeleid rondom AI als adequaat beoordeelt. Veel partijen binnen de overheid hebben momenteel te maken met AI-gerelateerde vraagstukken en weten daar maar beperkt raad mee. Een aantal van hen staat bij het zoeken naar antwoorden weliswaar met elkaar in contact, maar van een structureel gecoördineerde aanpak is geen sprake.

De afgelopen jaren hebben diverse inventarisaties van AI-toepassingen binnen en buiten de (centrale en decentrale) overheid het licht gezien, evenals verkenningen en adviezen over de relatie tot de daarbij betrokken publieke waarden. Veel van deze studies landen echter in een versnipperd landschap van betrokken en verantwoordelijke instanties. Een systeemtechnologie als AI vereist dat dit proces van kennisverwerving een permanent en gestructureerd karakter krijgt en dat informatie breed wordt gedeeld en besproken. Dit is noodzakelijk om zicht te krijgen op wat er gebeurt rond de inbedding van AI en de inrichtingsvraagstukken die deze voor de overheid met zich meebrengt.

Als eerstvolgende stap in de opbouw van een beleidsinfrastructuur bepleit de WRR daarom een coördinatiecentrum voor AI. Het gaat ons daarbij niet zozeer om de organisatie en naamgeving als zodanig, maar veeleer om de functies die het dient te vervullen.

Mogelijke functies van een AI-coördinatiecentrum

- *Platformfunctie* – Het centrum kan contact faciliteren tussen overheidsorganisaties op het niveau van beleid, uitvoering en evaluatie. Het kan bovendien fungeren als contactpunt voor internationale organisaties, met de EU en de voornoemde European Artificial Intelligence Board voorop.
- *Kennisfunctie* – Het centrum kan bijeenbrengen welke initiatieven en trajecten er lopen op het gebied van AI binnen en buiten de overheid. Dit kan bijvoorbeeld invulling krijgen in een jaarlijkse te verschijnen monitor over de staat van AI in Nederland, zoals ons land bijvoorbeeld ook een monitor Brede Welvaart kent. Op basis hiervan kunnen prioriteiten worden gesteld voor bijvoorbeeld de noodzaak van training op bepaalde domeinen of het identificeren van knelpunten. De inhoud van een dergelijke monitor kan jaarlijks – zoals bij de monitor Brede Welvaart – met de leden van de Tweede Kamer worden besproken.
- *Faciliterende functie* – Verder kan het coördinatiecentrum een belangrijke rol spelen bij de andere aanbevelingen om AI in de samenleving in te bedden. Zo kan het op hoofdlijnen de ontwikkeling van een AI-brevet voor overheidspersoneel ter hand nemen en *better intelligence* verzamelen ten behoeve van reguleringsvraagstukken.
- *Positionering* – Het coördinatiecentrum zou onafhankelijk moeten zijn, maar om kennisdeling, samenwerking en samenhangend beleid te stimuleren binnen de rijksoverheid ondergebracht moeten worden, bij een of meerdere ministeries. Tegelijkertijd is het belangrijk dat dit centrum wordt gevoed door kennis van buiten, zoals de wetenschap en het bedrijfsleven. Om hieraan tegemoet te komen valt te denken aan een externe AI-Raad van prominente experts, die periodiek bijeenkomen om het AI-coördinatiecentrum en de bredere overheid te informeren en adviseren.

Met de verdere uitwerking van de geschetste functies kan het door ons voorgestelde coördinatiecentrum aan beleidsdirecties, toezichthouders en uitvoeringsorganisaties een structuur bieden waardoor ze regelmatig en rond uiteenlopende kwesties met elkaar in contact treden. Omdat over domeinen heen – van zorg tot onderwijs en landbouw – vergelijkbare vragen leven, kunnen zij van elkaar leren hoe hiermee om te gaan. Een coördinatiecentrum kan bovendien focus aanbrengen in de voor de overheid relevante vraagstukken, kansen en risico's rondom AI. Alhoewel het coördinatiecentrum niet noodzakelijkerwijs gericht dient te zijn op centraal beleid – het brengt in eerste instantie

slechts bij elkaar wat op tal van plekken binnen de overheid wordt gesignaleerd, geprobeerd en gebruikt –, kan het een belangrijke coördinerende en faciliterende rol vervullen bij het opstellen van de bredere wetgevingsagenda die de WRR in aanbeveling 7 bepleit. De ervaringen hiermee kunnen in een volgende fase de basis vormen voor het faciliteren van beleidsvoorbereiding en wellicht ook beleidsbepaling en -uitvoering.

Hoewel het door ons voorgestelde coördinatiecentrum vooralsnog geen bevoegdheid heeft om beleid te maken, vervult het een cruciale rol bij de totstandkoming van beleid, moet er werk worden gemaakt van zijn bevindingen en staat het nauw in verband met politiek en bestuur. Het is daarom belangrijk dat het centrum een politieke verankering kent, zodat er als dat nodig is snel beleid kan worden gemaakt en daartoe politieke afstemming en sturing voorhanden is. Eerder al pleitte de Cyber Security Raad voor het instellen van een ministeriële onderraad waar cyberweerbaarheid integraal aan de orde zou moeten komen.[926] Aanhakend bij dit voorstel beveelt de WRR de regering aan een ministeriële onderraad in te stellen waar zwaarwegende kwesties rondom digitalisering die om een integrale afstemming vragen, aan de orde komen. Daartoe kunnen kwesties rondom cyberweerbaarheid maar zeker ook AI behoren. De WRR bepleit de instelling van een dergelijke ministeriële onderraad tevens vanuit de overtuiging dat digitalisering anno 2021 niet anders dan ook een politieke kwestie kan zijn.[927]

926 Cyber Security Raad 2021.
927 En daarmee een aanvulling op de recentelijk door de Tweede Kamer ingestelde Vaste Kamercommissie voor Digitale Zaken.

Figuur 9.3 Aanbevelingen per opgave voor de maatschappelijke inbedding van AI

9.8 Tot slot – De verbrandingsmotor van de eenentwintigste eeuw

De auto is tegenwoordig een vanzelfsprekend onderdeel van onze leefomgeving en dagelijks functioneren. Het is dan ook moeilijk voorstelbaar hoe revolutionair het voertuig ooit was. Laten we ons de situatie van precies honderd jaar geleden proberen voor te stellen. De verbrandingsmotor bestond in 1921 al een tijd, maar pas een paar jaar eerder had Henry Ford bewezen auto's massaal te kunnen produceren. Mensen wisten niet goed waar ze mee te maken hadden en spraken daarom van een *'horseless carriage'*. Ook was er scepsis over het nut van auto's. Dat was niet zo gek, want die hadden nog allerlei mankementen. Voor veel doeleinden bleven paarden beter geschikt. Er was bovendien nog geen goed wegennet waarop de auto optimaal kon functioneren.

Mettertijd zou die auto echter het aangezicht van stad en land en onze manier van leven flink gaan veranderen. Er volgde een 'strijd om de straat' waarin fietsers, voetgangers en zij die geen auto konden betalen, uiteindelijk van delen van de weg geweerd zouden worden. Maar de ontwikkeling droeg ook bij aan een ander besef van vrijheid en individualiteit. Door dergelijke grote veranderingen beïnvloedde de auto de inrichting van de hele samenleving en dat vroeg om regels, maatregelen en nieuwe instanties. Daarnaast vroeg de auto om een perspectief op de bredere inrichtingsvraagstukken. Zowel de individuele maatregelen als het bredere perspectief waren ook nodig om allerlei tweede-orde-effecten zoals vervuiling en gevaarzetting aan te pakken. Autobedrijven werden symbolen van vooruitgang en de nationale trots van landen. Tijdens de Tweede Wereldoorlog zou de verbrandingsmotor in allerlei voertuigen zijn stempel drukken op de oorlogsvoering.

Die ontwikkelingen waren in 1921 nog onmogelijk te voorzien. Er valt achteraf ook geen simpel antwoord te geven op de vraag hoe de auto de samenleving heeft veranderd en of dat iets goeds of iets slechts was. Zeker is wel dat het een spannend en langdurig proces was, en nog steeds is, om dit voertuig in de samenleving in te bedden.

Over honderd jaar zullen wij AI net zo vanzelfsprekend vinden als de auto nu is. In wat voor wereld we dan leven, kunnen wij ons nu niet voorstellen. Omgekeerd zal het ook moeilijk zijn om op dat moment een eeuw terug te kijken naar wat daar allemaal aan vooraf is gegaan, om ons dan voor te stellen hoe AI ooit in het lab begon en zich vervolgens gedurende decennia in de samenleving heeft verspreid. Wij staan nu aan de vooravond van dat proces. Met de opgaven die we in dit rapport onderscheiden en de bijbehorende aanbevelingen aan de regering beoogt de WRR eraan bij te dragen dat wij het spannende pad dat voor ons ligt, zo goed mogelijk belopen.

Aanbevelingen

Demystificatie

1. Maak leren over AI en de toepassing daarvan tot een expliciet doel bij het handelen door de overheid.
2. Stimuleer als overheid de ontwikkeling van AI-wijsheid bij het brede publiek, te beginnen met het opzetten van algoritmeregisters.

Contextualisering

3. Kies expliciet voor een Nederlandse AI-identiteit en onderzoek waar in de betreffende domeinen aanpassingen aan de technische omgeving nodig zijn.
4. Versterk de vaardigheden en het kritisch vermogen van individuen die met AI-systemen werken en ontwikkel daarvoor een stelsel van opleiding en certificering.

Engagement

5. Versterk de capaciteit van maatschappelijke organisaties om hun werk te verbreden naar het digitale domein, in het bijzonder met betrekking tot AI.
6. Draag zorg voor een goede terugkoppeling tussen de ontwikkelaar van AI, de gebruiker ervan en de personen die er in de praktijk de consequenties van ondervinden.

Regulering

7. Koppel de regulering van AI aan een discussie over de inrichting van de digitale leefomgeving en stel een brede wetgevingsagenda op.
8. Stuur via wetgeving actief op ontwikkelingen rondom surveillance en dataverzameling, de scheve verhouding tussen publiek en privaat in het digitale domein en machtsconcentratie.

Positionering

9. Versterk het Nederlandse verdienvermogen met een 'AI-diplomatie' die gericht is op internationale samenwerkingsverbanden, in het bijzonder binnen de EU.
10. Weet je als land ook in het AI-tijdperk te verdedigen: versterk daarom de Nederlandse capaciteiten tegen de groeiende 'informatieoorlog' en de export van digitale dictatuur.

Slotaanbeveling

Bouw een beleidsinfrastructuur voor AI op, te beginnen met een AI-coördinatiecentrum voorzien van politieke verankering middels een ministeriële onderraad.

Bijlage: Voorbeelden van AI-toepassingen in Nederland

Rijksoverheid	▪ Anticiperen op onderhoud aan infrastructuur ▪ In kaart brengen welke burgers hulp nodig hebben bij werkeloosheid ▪ Automatisering van contact met burgers ▪ Voorspellen van kans op fraude via risico-indicaties ▪ Risico-gestuurde inspecties ▪ Beoordeling van documenten op volledigheid ▪ Beveiliging van mailverkeer ▪ Geautomatiseerde bediening van bruggen en waterwerken ▪ Vertaling van productinformatie van ingevoerde goederen
Gemeenten	▪ Categoriseren van probleemmeldingen over de openbare ruimte ▪ Reguleren van verkeerslichten ten gunste van hulpdiensten of fietsers ▪ Geautomatiseerde risico-indicatie voor bijstandsfraude ▪ Crowd management ▪ Toegang tot milieuzones via kentekendetectie
Politie	▪ Anticiperen op patroonmatige criminaliteit ▪ Geautomatiseerde hulp bij online aangiften ▪ Inschatten van de kansrijkheid van cold cases ▪ Matchen van foto's van verdachten met een bestaande fotodatabase
Onderwijs	▪ Digitale surveillance tijdens toetsmomenten ▪ Adaptieve leermiddelen voor basisschoolleerlingen
Zorg	▪ Ondersteuning bij triage op de IC ▪ Beoordeling van medisch beeldmateriaal ▪ Automatische verslaglegging van consulten ▪ Ondersteuning van diagnostiek

Financiële sector	▪ Voorspellen van prijstrends voor beleggers ▪ Beoordelen van kredietwaardigheid van klanten ▪ Selectief toekennen van kortingen op verzekeringspremies
Agrifood	▪ Monitoring van dierenwelzijn in stallen ▪ Kwaliteitsinspectie van gewassen ▪ 'Just-in-time'-onderhoud van machines ▪ Beoordelen bodemkwaliteit middels satellietdata
Detailhandel	▪ Voorspellen van producthoudbaarheid ▪ Dynamisch prijzen van producten ▪ Personaliseren van marketing ▪ Voorspellen van koopgedrag
Media	▪ Geautomatiseerde productie van nieuwsartikelen ▪ *Microtargetting* via sociale media ▪ Gepersonaliseerd aanbod van content
Rechtspraak	▪ Doorzoeken van jurisprudentie ▪ Geautomatiseerde online geschillenbeslechting
Werkvloer	▪ Voorspellen van matches bij sollicitaties ▪ Analyseren van de prestaties van werknemers ▪ Geautomatiseerde routing voor chauffeurs

Gesprokenpersonen

Vermelding organisatie ten tijde van gesprek

E. (Emile) Aarts, Tilburg University, Nederlandse AI Coalitie
R. (Ron) Augustus, SURF
R. (Robert) Baris, De Belastingdienst
A. (Arie) van Bellen, ECP | Platform voor de informatiesamenleving
S. (Salima) Benhamou, France Strategie
E. (Ellen) Berends, Onderzoeksraad voor de Veiligheid
S. (Siri) Berends, Set Up
F. J. (Floris) Bex, Universiteit Utrecht
P. (Patrick) Blankers, Ericsson Nederland
F. (Floris) den Boer, PIANOo
V. (Vincent) Böhre, Privacy First
C. (Christien) Bok, SURF
R. (Romain) Bonenfant, Ministère de l'Économie, des Finances et de la Relance, France
H. (Hans) Bos, Microsoft
J. A. (Jan) van den Bos, Inspectieraad
M. (Mirèl) ter Braak, Autoriteit Financiële Markten
B. (Barteld) Braaksma, CBS
A. (Alain) Bravo, L'Académie des Technologies, France
J. (Joël) Buiter, Inspectie Gezondheidszorg en Jeugd
M. (Mirjam) van Burgel, Agentschap Telecom
J. (Joost) van der Burgt, De Nederlandsche Bank
F. (Frits) Bussemaker, I-Partnerschap Rijk – Hoger Onderwijs, Institute for Accountability in the Digital Age
P. (Pierluigi) Casale, TomTom
M. (Madeleine) de Cock Buning, Universiteit Utrecht
S. (Stephanie) Combes, Ministère de la Santé, France
K. (Kate) Crawford, AI Now Institute
R. (Roxane) Daniëls, Vereniging van Nederlandse Gemeenten
P. (Petra) Delsing, Ministerie van Infrastructuur en Waterstaat
S. (Stijn) van Deursen, Universiteit Utrecht
M. (Marloes) Dignum, Ministerie van Infrastructuur en Waterstaat
M. V. (Virginia) Dignum, Umeå Universitet
J. F. T. M. (José) van Dijck, Universiteit Utrecht
K. H. D. M. (Klaas) Dijkhoff, VVD
R. I. J. (Roel) Dobbe, AI Now Institute, TU Delft
P. (Pedro) Domingos, University of Washington
L. (Louis) Dubertet, National Academy of Technologies of France
N. (Nathan) Ducastel, Vereniging van Nederlandse Gemeenten

M. (Myrte) Dujardin, Ministerie van Sociale Zaken en Werkgelegenheid
B. M. A. (Marlies) van Eck, Universiteit Leiden
A. (Aik) van Eemeren, Gemeente Amsterdam
Q. (Quirine) Eijkman, College voor de Rechten van de Mens
Q. C. (Rinie) van Est, Rathenau Instituut
B.J. (Bart Jan) van Ettekoven, Raad van State
T. (Thomas) Faber, Ministerie van Economische Zaken
G. (Gerard) Feitsma, Agentschap Telecom
B. (Bas) Filippini, Privacy First
G.J. (Gert-Jan) Fonk, Ministerie van Landbouw, Natuur en Voedselkwaliteit
P. H. A. (Paul) Frissen, Universiteit Tilburg
J. D. C. (Jacobine) Geel, College voor de Rechten van de Mens
Y. (Yannick) van Gelder, Wageningen University & Research
S. (Sennay) Ghebreab, Universiteit van Amsterdam
E. C. (Corine) van Ginkel, Wetenschappelijk Onderzoek- en Documentatiecentrum
P. (Peter) Goeijers, Ericsson Nederland
P. (Peter) Gouw, Ministerie van Volksgezondheid, Welzijn en Sport
J. (Jochem) de Groot, Microsoft
M. (Marion) Gust, Ministère de la Transition écologique et solidaire, France
T. (Tom) van de Haar, Ministerie van Sociale Zaken en Werkgelegenheid
H. J. (Hugo) van Haastert, Ministerie van Volksgezondheid, Welzijn en Sport
P. (Peter) Hagendoorn, The Fluid Society
G. (Gry) Hasselbalch, DataEthics
M. (Martin) Heijnsbroek, MIcompany
R. (Rik) Helwegen, Universiteit van Amsterdam
S. (Steven) Hillebrink, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
J. (Justin) Hoegen Dijkhof, College voor de Rechten van de Mens
R. (Ronald) van den Hoogen, Rijksacademie voor Digitalisering en Informatisering Overheid
F. (Floris) Hoogenboom, Schiphol
H. H. (Holger) Hoos, Universiteit Leiden
G. (Gerald) Hopster, Autoriteit Persoonsgegevens
N. (Noor) Huiboom, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
B. (Bas) van Hulst, DeepVision
J. P. (Joost) van Iersel, European Policy Centre
I. (Ivana) Isgum, Universiteit van Amsterdam
M. (Menno) Israël, Autoriteit Consument en Markt
H. (Hans) van de Jagt, Ministerie van Defensie
C. (Caspar) de Jonge, Ministerie van Infrastructuur en Waterstaat
E. (Erik) Jonker, Algemene Rekenkamer

C. M. (Catholijn) Jonker, TU Delft
R. (Rina) Joosten-Rabou, Seedlink
M. (Merel) Kampers, Ministerie van Sociale Zaken en Werkgelegenheid
C. (Chandro) Kandiah, Ministerie van Landbouw, Natuur en Voedselkwaliteit
D. (Daan) Keijser, Douane
M. C. G. (Mona) Keijzer, Ministerie van Economische Zaken en Klimaat
K. (Kees) van der Klauw, Nederlandse AI Coalitie
M. H. (Meine Henk) Klijnsma, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
R. (Rogier) Klimbie, Considerati
V. (Victor) Klos, Autoriteit Persoonsgegevens
L. (Linda) Kool, Rathenau Instituut
M. (Mauritz) Kop, AIRecht, Stanford University
A. (Antoine) de Kort, Rijksdienst Wegverkeer
W. (Willem) Korteweg, Leading Edge Forum
K. (Katja) van Kranenburg, CMS, Nederlandse AI Coalitie
F. (Floris) Kreiken, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
J. (Johan) Krijgsman, Inspectie Gezondheidszorg en Jeugd
W. (Wouter) Kroese, Pacmed
A. H. (Bert) Kroese, CBS
R. L. (Inald) Lagendijk, Technische Universiteit Delft
T. (Tom) Leenders, Ministerie van Financiën
M. H. G. (Michel) van Leeuwen, Ministerie van Justitie en Veiligheid
A. (Anja) Lelieveld, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
O. (Olivia) Lin, Ministerie van Buitenlandse Zaken
F. (Frans) Lips, Ministerie van Landbouw, Natuur en Voedselkwaliteit
G. (Gilles) de Margerie, France Stratégie
A. (Alexander) Melchior, Ministerie van Landbouw, Natuur en Voedselkwaliteit
J. (Jeroen) van Mierlo, Ministerie van Onderwijs, Cultuur en Wetenschap
I. (Inge) Molenaar, Radboud Universiteit
B. (Bennie) Mols, Wetenschapsjournalist
S. (Selwyn) Moons, PricewaterhouseCoopers
S. (Sander) Mul, Ministerie van Justitie en Veiligheid
Y. (Yvette) Mulder, NEN
C. C. J. (Catelijne) Muller, ALLAI
J. (Jasper) Nagtegaal, Agentschap Telecom
E. (Edwin) Nas, Ministerie van Infrastructuur en Waterstaat
R. (Rob) Nijman, IBM
C. (Carine) van Oosteren, Sociaal-Economische Raad
C. (Cees) Oudshoorn, VNO-NCW
B. (Bertrand) Pailhes, CNIL

G. (Geert) Pater, Rijksdienst Wegverkeer
M. (Miranda) Pirkovski, Algemene Rekenkamer
T. (Theo) van de Plas, Nationale Politie
A. (Addy) Polet, Ministerie van Onderwijs, Cultuur en Wetenschap
M. (Marieke) van Putten, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
A. (Ardaan) van Ravenzwaaij, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
B. (Bram) de Rijk, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
M. (Maarten) de Rijke, Universiteit van Amsterdam
D. (Dirk) van Roode, NL Digital
L. (Lennart) Salemink, Ministerie van Infrastructuur en Waterstaat
T. (Tim) Salimans, Google Brain
J. (Johan) Schot, Universiteit Utrecht
G. (Gijs) van Schouwenburg, Ministerie van Infrastructuur en Waterstaat
O. (Olof) Schuring, Ministerie van Justitie en Veiligheid
C. (Cecile) Schut, Autoriteit Persoonsgegevens
A. (Arnold) Smeulders, Universiteit van Amsterdam
C. (Cees) Snoek, Universiteit van Amsterdam
J. (Just) Stam, Ministerie van Justitie en Veiligheid
M. (Mildo) van Staden, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
M. R. (Maarten) van Steen, Nederlandse AI Coalitie
M. (Michiel) Steltman, Digitale Infrastructuur Nederland
B. (Bart) Stuut, Autoriteit Consument en Markt
J. T. (Jonathan) Taplin, University of Southern California
L. (Linnet) Taylor, Universiteit van Tilburg
L. E. M. (Linda) Terlouw, DeepVision
M. (Marthe) Tholen, De Belastingdienst
B. (Benjamin) Timmermans, IBM
B. (Bert) Timmermans, Ministerie van Infrastructuur en Waterstaat
A. (Amarens) Veeneman, Bureau Tweede Kamer
M. (Maarten) Veltman, Douane
P. P. C. C. (Peter-Paul) Verbeek, Universiteit Twente
P. (Peter) Vermeulen, Ministerie van Sociale Zaken en Werkgelegenheid
C. D. (Corinne) Vigreux, TomTom, CODAM
F. W. (Focco) Vijselaar, Ministerie van Economische Zaken en Klimaat
I. (Iris) Vissers, Ministerie van Financiën
J. (Jasper) van Vliet, Inspectie Leefomgeving en Transport
B. (Bart) Voorn, Ahold Delhaize
M. (Michael) Vos, Microsoft
C. H. (Claes) de Vreese, Universiteit van Amsterdam
R. (René) Vroom, Agentschap Telecom

L. (Lauren) Waardenburg, Vrije Universiteit
M. (Marieke) van Wallenburg, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
H. (Hans) Wanders, Algemene Bestuursdienst
S. (Sandra) van der Weide, Ministerie van Economische Zaken en Klimaat
I. (Inge) Welbergen, Ministerie van Onderwijs, Cultuur en Wetenschap
M. (Max) Welling, Universiteit van Amsterdam
K. (Klaas) Werkhorst, Ministerie van Justitie en Veiligheid
I. (Inge) Wertwijn, De Belastingdienst
R. (Ronald) Westerhof, Gemeente Apeldoorn
R. (Richard) Weurding, Verbond van Verzekeraars
K. (Koen) Wienk, Nationale Voedsel en Warenautoriteit
J. (Joost) Witteman, SEO
A. (Aleid) Wolfsen, Autoriteit Persoonsgegevens
S. M. E. (Sally) Wyatt, Maastricht University
R. (Rolf) Zeldenrust, PIANOo
B. (Berrie) Zielman, Algemene Rekenkamer
R. F. B. (Reinier) van Zutphen, De Nationale Ombudsman
B. (Bart) Zwartjes, Autoriteit Financiële Markten
R. (Richard) van Zwol, Raad van State
G. (Guus) van Zwoll, Ministerie van Buitenlandse Zaken

Begrippenlijst

AI-effect Het effect van de ontwikkeling van computervaardigheden op wat wij beschouwen als menselijke intelligentie. Zodra een computer een bepaalde vaardigheid beheerst, zijn mensen geneigd dat niet als intelligent gedrag maar als simpele calculatie te zien.

AI-winter Een periode met relatief weinig wetenschappelijke vooruitgang op het gebied van AI. Tot nu toe hebben er twee AI-winters plaatsgevonden: de eerste vanaf 1969 tot aan 1982 en de tweede vanaf het eind van de jaren tachtig tot de jaren negentig.

AI-zomer Een periode met relatief veel wetenschappelijke vooruitgang en activiteit op het gebied van AI. Momenteel bevinden wij ons in een AI-zomer. Deze periode begon rond 2012, toen de wetenschappelijke doorbraken binnen de benadering van neurale netwerken zorgden voor een explosie aan activiteit in het veld van AI.

Algoritme Een gespecificeerde instructie om een probleem op te lossen of een berekening uit te voeren.

AlphaGo Een computerprogramma dat het bordspel *Go* speelt. Het programma werd ontwikkeld door het Britse onderzoekslab DeepMind, dat in 2014 werd overgenomen door Google. In 2016 versloeg AlphaGo de wereldkampioen Lee Sedol.

Artificial General Intelligence (AGI) Artificiële Intelligentie die menselijke intelligentie op alle gebieden kan evenaren. Dit wordt ook wel *strong* of *full AI* genoemd.

Artificial Super Intelligence (ASI) Artificiële Intelligentie die op alle gebieden superieur is aan menselijke intelligentie.

Artificiële Intelligentie Systemen die intelligent gedrag vertonen door hun omgeving te analyseren en acties te ondernemen – met enige graad van autonomie – om specifieke doelen te bereiken. Hiermee sluit dit rapport aan bij de definitie van de AI HLEG van de Europese Commissie.

Artificiële Neurale Netwerken (ANN's, zie ook connectionisme) Een benadering binnen het veld van AI dat gebruikmaakt van artificiële neurale netwerken om de werking van neuronen in het menselijk brein te simuleren. Deze netwerken worden met grote hoeveelheden data gevoed waaruit het systeem zelf patronen moet destilleren. Bij deze benadering worden vooraf dus geen door mensen opgestelde regels meegegeven.

Automation bias Een psychologisch mechanisme waarbij mensen blind vertrouwen op technologie en zonder nadenken de suggesties van een computer volgen, ook als die incorrect zijn of tegen gezond verstand ingaan.

Backpropagation Een algoritme dat wordt gebruikt om de patroonherkenning van artificiële neurale netwerken (ANN's) te verbeteren. Het algoritme kijkt hier naar de uitkomsten van de 'outputlaag' en traceert welke informatie uit de verborgen lagen (onder de outputlaag) komt. Vervolgens kan het algoritme de individuele units identificeren die aangepast moeten worden om het algoritme beter te laten functioneren.

Blockchain Een decentraal systeem waarin transacties onveranderlijk en openbaar worden geregistreerd. De term 'blockchain' verwijst naar de structuur van de database. Een nieuwe transactie vormt namelijk een 'blok' met informatie over de nieuwe transactie en informatie over de voorgaande transactie. Als de transactie wordt goedgekeurd, sluit het blok zich aan op de andere blokken waardoor de blokken samen een 'informatieketen' vormen.

Central processing unit (CPU) Een processor (stuk hardware) die de basisinstructies van de programmacode ontvangt, aanstuurt en uitvoert.

Civiele technologie Technologie die door bedrijven is ontwikkeld en niet (of in mindere mate) door de overheid of onafhankelijke wetenschappers.

Closure Een moment in de ontwikkeling van een technologie waarop een bepaald ontwerp of gebruik de norm wordt. De controverse rondom de technologie verdwijnt dan vaak.

Collingridge-dilemma

Dit dilemma gaat over het moment van reguleren en de onzekerheid die daarmee gepaard gaat. In een vroege fase van nieuwe technologie is het makkelijk om kaders op te stellen, maar is het vaak nog onduidelijk aan wat voor regulering behoefte is. Daarenboven ontbreekt de noodzaak voor regulering omdat de effecten nog onbekend zijn. In een latere fase wordt duidelijker waar regels nodig zijn, maar is reguleren een stuk lastiger en kostbaarder doordat al praktijken en belangen zijn ontstaan.

Computer vision

AI die gericht is op het waarnemen, analyseren en interpreteren van visuele informatie zoals foto's, video's en de fysieke omgeving. Een van de bekendste toepassingen van computer vision is gezichtsherkenning.

Connectionisme

Een benadering binnen het veld van AI dat gebruikmaakt van artificiële neurale netwerken om de werking van neuronen in het menselijk brein te simuleren (zie ook Artificiële Neurale Netwerken). Deze benadering is naast de symbolische benadering een van de twee hoofdstromingen binnen AI.

Criminaliteit Anticipatie Systeem (CAS)

Een voorspelsysteem van de politie dat patroonmatig criminaliteit in kaart brengt en voorspelt waar en wanneer de kans op een voorval het grootst is. Agenten gebruiken deze informatie om op criminaliteit te kunnen anticiperen door bijvoorbeeld extra te surveilleren in een bepaalde regio.

Data

Gegevens die door een computer kunnen worden opgeslagen. Meestal zijn dit individuele gegevens zonder nadere toelichting, waardoor ze in een bepaalde context of verzameling van gegevens geplaatst moeten worden om geïnterpreteerd te kunnen worden.

Deep Blue

Een schaakcomputer ontwikkeld door IBM. In 1997 won Deep Blue van de schaakgrootmeester Garry Kasparov.

Deepfake

Een door AI gegenereerd beeld of geluidsfragment met de suggestie dat het gaat om ongemanipuleerde content. Deepfakes kunnen dusdanig realistisch lijken dat het moeilijk is om ze van echt te onderscheiden.

Deep learning (DL) Een vorm van machine learning waarbij de werking van neuronen in het menselijk brein wordt gesimuleerd. Deep learning maakt gebruik van meerlaagse netwerken, vandaar de term 'deep'.

Dual use-technologie Technologie die zowel voor civiele als militaire doeleinden kan worden ingezet.

Enveloping Het aanpassen van een omgeving zodat een technologie daarbinnen (beter) kan functioneren.

Expertsystemen AI-systemen die vooraf geprogrammeerde regels volgen, gebaseerd op de kennis van experts. Expertsystemen vallen onder regelgebaseerde AI.

Federated learning Een vorm van machine learning waarbij algoritmen worden verbeterd door hun parameters aan te passen aan de parameters van andere datasets, zonder de data in deze datasets te combineren. De dataset hoeft in dit geval niet te worden opgenomen in een centrale server. Het algoritme gaat als het ware naar de data toe in plaats van andersom.

General Purpose Technology (GPT) Technologieën die niet één beperkte toepassing maar een generiek karakter hebben en daarom in talloze vormen kunnen worden toegepast voor uiteenlopende doeleinden. Voorbeelden van eerdere *general purpose technologies* zijn de verbrandingsmotor en elektriciteit.

Generative Adversarial Networks (GAN's) Een techniek binnen AI waarbij verschillende algoritmen worden gebruikt om elkaar te verbeteren. Het ene algoritme genereert iets nieuws (zoals een afbeelding), waarna het andere algoritme probeert te detecteren of het gefabriceerd of authentiek is. Het eerste algoritme gaat door met genereren totdat het tweede algoritme ervan overtuigd is dat product van het eerste algoritme authentiek is.

Geo-economie Het domein dat zich vormt op het snijvlak van economie en verdienvermogen enerzijds en geopolitiek en veiligheid anderzijds.

GPT-3 (Generative Pre-trained Transformer 3)	Een taalverwerkingssoftware die met relatief weinig input natuurlijke teksten kan genereren.
Graphic Processing Units (GPU's)	Processors (hardware) die doorgaans worden gebruikt voor het verwerken van complexe afbeeldingen en grafische data. Door hun grote rekenkracht zijn GPU's ook geschikt voor de zware berekeningen die nodig zijn voor geavanceerde AI.
Human-in-the-loop	Een vorm van interactie tussen mens en machine waarbij een AI-systeem in een proces betrokken is terwijl de verantwoordelijkheid van de beslissing bij de mens ligt.
Human-on-the-loop	Een vorm van interactie tussen mens en machine waarbij het AI-systeem zelfstandig beslissingen kan nemen zonder dat er een mens aan te pas komt. Wel heeft een mens inzicht in het proces en is die in staat om in te grijpen en wijzigingen aan te brengen.
Human-out-of-the-loop	Een vorm van interactie tussen mens en machine waarbij geen mens meer betrokken is en het AI-systeem volledig autonoom handelt.
Internet of Things (IoT)	Het netwerk van digitale verbindingen tussen objecten en apparaten in de fysieke omgeving via sensoren en het internet, zodat daartussen onderling informatie kan worden uitgewisseld.
Logische AI	Zie Symbolische AI.
Luddieten	Benaming voor de Engelse fabrieksmedewerkers die in de jaren tien van de negentiende eeuw in opstand kwamen tegen de mechanisering van arbeid.
Machine learning (ML)	Een bepaald type toepassing van AI die, door patronen in datasets te zoeken, voorspellende analyses kan afleiden.
Microtargeting	Het zo nauwkeurig mogelijk afstemmen van advertenties op de interesses en gevoeligheden van individuele gebruikers voor commerciële of politieke doeleinden.

Model	Een formele omschrijving van een systeem, proces of vergelijking dat wordt gebruikt om een complex onderwerp te vereenvoudigen.
Natural Language Processing (NLP)	AI die is gericht op het lezen, analyseren en genereren van menselijke taal. Het doel is hier om algoritmen de 'natuurlijke' taal zo goed mogelijk te laten begrijpen, zodat ze taken kunnen uitvoeren waarvoor het vereist is tekst te interpreteren.
Paradox van Moravec	Het fenomeen dat bepaalde zaken die voor mensen moeilijk zijn, voor computers makkelijk zijn en vice versa.
Productiviteits-paradox	Het fenomeen dat, hoewel de verwachtingen rondom nieuwe technologieën vaak hooggespannen zijn, de daadwerkelijke invloed van nieuwe technologieën op de productiviteit van de economie op de korte termijn doorgaans tegenvalt.
Proxy (proxies)	Een proxy is een datapunt dat indicatief is voor andere gegevens. Op basis van proxies kan dus informatie worden gereconstrueerd die niet direct wordt gemeten of verzameld. Zo kan taalgebruik bijvoorbeeld een proxy zijn voor geslacht of kan een postcode een proxy zijn voor etniciteit.
Quantum computing	Technologie die werkt met *quantum bits* of *qubits*. Quantum bits of qubits kunnen meerdere staten tegelijkertijd aannemen, waardoor het aantal mogelijke berekeningen veel hoger is dan bij traditionele computers die werken met bits in een binaire logica.
Regelgebaseerde AI	Zie Symbolische AI.
Reinforcement learning	Een vorm van machine learning waarbij het algoritme wordt getraind om bepaalde strategieën te volgen via een systeem van positieve en negatieve feedback.
Soft law	Soft law bestaat uit normen, codes en aanbevelingen, die uit hun aard weinig dwingende of verbindende kracht hebben. Wel kan van soft law een zekere *opinio iuris* uitgaan.

Speech recognition Het domein binnen AI dat zich bezighoudt met de waarneming, analyse en interpretatie van gesproken menselijke taal. Hier worden algoritmes gebruikt om zinnen en woorden te ontwaren uit gesproken taal en deze om te zetten in een tekstformat zodat dat kan worden geanalyseerd.

Strong AI Zie Artificial General Intelligence (AGI).

Supervised learning Een vorm van machine learning waarbij een systeem wordt gevoed met door mensen gelabelde data.

Symbolische AI Een benadering van AI die is gebaseerd op logische regels en formules. Hierbij worden bijvoorbeeld 'als-dan'-regels opgesteld waarmee het programma redeneert over de data. Dit wordt ook wel 'logische' of 'regelgebaseerde' AI genoemd. Symbolische AI is naast het connectionisme een van de twee hoofdstromingen binnen het veld van AI.

Systeem Risico Indicatie (SyRI) Systeem Risico Indicatie (SyRI) is een wettelijk instrument van de Nederlandse overheid voor de bestrijding van fraude op bijvoorbeeld het terrein van uitkeringen, toeslagen en belastingen. SyRI werd gebruikt door enkele gemeenten totdat de rechter in 2020 het systeem als onrechtmatig beoordeelde omdat het in strijd is met het recht op respect voor het privéleven.

Techno-chauvinisme De overtuiging dat technologie een oplossing kan bieden voor elk probleem.

Techno-determinisme/ technologisch determinisme Het idee dat technologie een autonome ontwikkeling kent waaraan de samenleving zich moet aanpassen.

Techno-solutionisme/ technologisch solutionisme De neiging om complexe sociale fenomenen om te dopen tot vraagstukken waarop technologie het antwoord is.

Tensor Processing Unit (TPU)	Een processor (stuk hardware) die speciaal is ontworpen voor applicaties van machine learning.
Turingtest	Een experiment bedacht door Alan Turing in 1950, waarbij een computer zich voordoet als een mens. De computer slaagt voor de test als een mens niet kan vaststellen of de antwoorden worden gegeven door een mens of een computer. Varianten van deze test worden gebruikt wanneer AI-systemen worden vergeleken met menselijke vermogens, zoals het gebruik van taal.
Unsupervised learning	Een vorm van machine learning waarbij het programma wordt gevoed met ongelabelde data, waaruit het algoritme zelf patronen moet destilleren.
Watson	Watson is een computerprogramma dat werd ontwikkeld door IBM om gesproken vragen te beantwoorden tijdens de spelshow *Jeopardy!*. Dat gebeurde door middel van grote hoeveelheden informatie uit een database. In 2011 versloeg het programma de regerende menselijke kampioenen.
Weak AI (ook wel narrow AI)	Artificiële intelligentie die is gericht op een specifieke vaardigheid zoals beeld- of spraakherkenning, vaak binnen een specifieke context.
Wet van Moore	Het waargenomen patroon dat het aantal transitoren op een chip grofweg elke twee jaar verdubbelt.

Sleutelbegrippen

AI-diplomatie	Een geïntegreerd internationaal beleid dat is gericht op samenwerkingsverbanden op het gebied van AI. Dit kan betrekking hebben op vijf domeinen: fundamenteel onderzoek, commerciële toepassingen, regulering, ethische richtlijnen en standaarden.
AI-identiteit	Een onderscheidend karakter van een land of regio op het gebied van AI. Een AI-identiteit kan volgen uit de specialisatie in een bepaald type AI of het innemen van een bepaalde rol binnen het internationale speelveld rondom AI.
AI-wijsheid	Een basis van kennis en competenties die nodig is om deel te kunnen nemen aan een samenleving waarin AI een belangrijke rol speelt.
Contextualisering (Opgave 2)	Contextualisering betreft de inbedding van technologie in het socio-technische ecosysteem waarbinnen ze moet functioneren. Het idee is dat een technologie pas gaat werken zodra de technologie enerzijds en de technische en sociale context anderzijds goed op elkaar zijn afgestemd.
Demystificatie (Opgave 1)	Demystificatie betreft het tegengaan van overspannen voorstellingen of onjuiste beelden van een technologie, om zo meer begrip te creëren voor wat de technologie in feite inhoudt en kan.
Digitale leefomgeving	De ruimtelijke ordening en inrichting van het speelveld waarin technologie gebruikt wordt. In de digitale leefomgeving wordt technologie, door middel van regels en instrumenten, idealiter geplaatst op plekken waar ze een optimale bijdrage kan leveren, en ingezet op een schaal en voor doelen die congruent zijn met de (Nederlandse) samenleving.
Engagement (Opgave 3)	Engagement betekent dat verschillende groepen in de samenleving bij de ontwikkeling van een technologie zijn betrokken, om ervoor te zorgen dat hun belangen een plek krijgen in dat proces.

Positionering (Opgave 5) Positionering gaat over de strategische activiteit van een land met betrekking tot een technologie in het internationale speelveld. Hierbij gaat het om de relatie tot andere landen, maar ook tot niet-statelijke actoren zoals bedrijven en (criminele) organisaties.

Regulering (Opgave 4) Regulering is het ontwikkelen van kaders om de ontwikkeling en het gebruik van een technologie in goede banen te leiden. Dat kan via wet- en regelgeving, normen en standaarden, zowel op nationaal als internationaal niveau.

Systeemtechnologie Een multifunctionele technologie die zich verweeft in het systeem van de samenleving en functioneert binnen een systeem van andere technologieën. Net als *general purpose technologies* zijn systeemtechnologieën alomtegenwoordig, leiden ze tot complementaire innovaties en kennen ze continue technische verbetering. Met de term 'systeemtechnologie' wordt het systeemkarakter en de systemische effecten ervan benadrukt. De maatschappelijke inbedding van een systeemtechnologie betreft een complex en langdurig proces van inpassing en aanpassing.

Afkortingen

AGI	Artificial General Intelligence
AI HLEG	High-Level Expert Group on AI
AIV	Adviesraad Internationale Vraagstukken
ALLAI	Alliance for Artificial Intelligence Netherlands
ANNS	Artificiële Neurale Netwerken
AVG	Algemene Verordening Gegevensbescherming
CAPTCHA	Completely Automated Public Turing-tests to tell Computers and Humans Apart
CCW	Convention on Certain Conventional Weapons
CEN	Comité Européen de Normalisation
CENELEC	Comité Européen de Normalisation Electrotechnique
CPU	Central Processing Unit
DARPA	Defense Advanced Research Projects Agency
DIN	Deutsche Institut für Normung
DIU	Defense Innovation Unit
DKE	Deutsche Kommission Elektrotechnik Elektronik Informationstechnik
DL	Deep learning
DMA	Digital Markets Act
DSA	Digital Service Act
GMO	Genetically Modified Organism
GOFAI	Good Old-Fashioned AI
GPT	General Purpose Technology
GPU	Graphic Processing Unit
HRM	Human Resource Management
ICAI	Innovation Center for Artificial Intelligence
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
ITU	International Telecommunication Union
JAIC	Joint Artificial Center
ML	Machine learning
MVP	Minimal Viable Product
NEN	Nederlands Normalisatie Instituut
NSCAI	National Security Commission on Artificial Intelligence
OESO	Organisatie voor Economische Samenwerking en Ontwikkeling
OSS	open source software
SAPAI	Strategisch Actieplan voor AI
TPU	Tensor Processing Unit
VN	Verenigde Naties

Literatuur

Aanhangsel Handelingen II 2017/2018, nr. 1645 (2018, 4 april) *Aanhangsel van de handelingen*. Beschikbaar op: https://zoek.officielebekendmakingen.nl/ah-tk-20172018-1645.pdf

Access Now (2020) *Europe's approach to artificial intelligence: How AI strategy is evolving,* Brussel: Access Now. Beschikbaar op: https://www.accessnow.org/cms/assets/uploads/2020/12/Europes-approach-to-AI-strategy-is-evolving.pdf

Ackerman, E. (2021) 'How Boston Dynamics Taught Its Robots to Dance', *IEEE Spectrum*, 7 januari 2021. Beschikbaar op: https://spectrum.ieee.org/automaton/robotics/humanoids/how-boston-dynamics-taught-its-robots-to-dance

Ad Hoc Expert Group (2020) *First Draft of the Recommendation of the Ethics of Artificial Intelligence*, Parijs: UNESCO. Beschikbaar op https://unesdoc.unesco.org/ark:/48223/pf0000373434

Afdeling Bestuursrechtspraak van de Raad van State (2017) ECLI:NL:RVS:2017:1259 (AERIUS I), uitspraak 17 mei 2017. Beschikbaar op: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RVS:2017:1259

Afdeling Bestuursrechtspraak van de Raad van State (2018) ECLI:NL:RVS:2018:2454 (AERIUS II), uitspraak 18 juli 2018. Beschikbaar op: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RVS:2018:2454

AFM en DNB (2019) *Artificiële Intelligentie in de verzekeringssector een verkenning*, Den Haag: Autoriteit Financiële Markten, De Nederlandsche Bank. Beschikbaar op: https://www.afm.nl/~/profmedia/files/rapporten/2019/afm-dnb-verkenning-ai-verzekeringssector.pdf?la=nl-nl

Agrawal, A., J. Gans en A. Goldfarb (2018) *Prediction machines: the simple economics of artificial intelligence*, Boston: Harvard Business Press.

Agrawal, A., J.Gans en A. Goldfarb (reds.) (2019) *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research, Chicago: University of Chicago Press.

Ahmed, S. (2019) 'Credit Cities and the Limits of the Social Credit System', pp. 55-61 in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, Montgomery: Air University Press.

AIV en CAVV (2015) *Autonome wapensystemen: de noodzaak van betekenisvolle menselijke controle*, Nr. 97 AIV/Nr. 26 CAVV, Den Haag: Adviesraad Internationale Vraagstukken en Commissie van Advies Inzake Volkenrechtelijke Vraagstukken. Beschikbaar op: https://www.adviesraadinternationalevraagstukken.nl/documenten/publicaties/2015/10/02/autonome-wapensystemen

Albert Heijn (2019) *Albert Heijn zet kunstmatige Intelligentie in tegen voedselverspilling*, Albert Heijn Nieuws, 20 mei 2019. Beschikbaar op:

https://nieuws.ah.nl/albert-heijn-zet-kunstmatige-intelligentie-in-tegen-voedselverspilling/

Algemene Rekenkamer (2021) *Aandacht voor algoritmes,* Den Haag: Algemene Rekenkamer.

Alliance for Artificial Intelligence (z.d.) *Introducing ALLAI.* Beschikbaar op: https://allai.nl/about-us/

Amnesty International (2020a) *We sense trouble: Automated discrimination and mass surveillance in predictive policing in the Netherlands*, Londen: Amnesty International. Beschikbaar op: https://www.amnesty.org/en/wp-content/uploads/2021/05/EUR3529712020ENGLISH.pdf

Amnesty International (2020b) *Out of control: Failing EU laws for digital surveillance export*, Londen: Amnesty International.

Amsden, A. (1989) *Asia's next giant: South Korea and late industrialization*, Oxford: Oxford University Press.

Amsterdam Algoritmeregister (z.d.) *Anderhalve meter monitor.* Beschikbaar op: https://algoritmeregister.amsterdam.nl/anderhalve-meter-monitor/

Andersen, R. (2020) 'The Panopticon is already here', *The Atlantic*, september 2020. Beschikbaar op: www.theatlantic.com/magazine/archive/2020/09/china-ai-surveillance/614197/

Apple, C. (2020) 'Instant Gratification: The history of Instagram', *Spokesman*, 24 juni 2020. Beschikbaar op: https://www.spokesman.com/stories/2020/jun/24/how-instagram-hit-one-billion-users/

Asselt, M. van, E. Voss en T. Fox (2010) 'Regulating technologioes and the uncertainty paradox', blz. 261-286 in M. Goodwin, E. Koops en R. Leenes (reds.) *Dimensions of technology regulation*, Nijmegen: Wolf Legal Publishers.

Association for Advancing Automation (2018) 'Why AI won't overtake the world, but is worth watching', *Industry Insights*, 25 januari 2018. Beschikbaar op: www.robotics.org/content-detail.cfm/Industrial-Robotics-Industry-Insights/Why-AI-Won-t-Overtake-the-World-but-Is-Worth-Watching/content_id/6979

Austin, E. (2019) 'Facebook liegt tegen Tweede Kamer over verkiezingsmanipulatie', *Bits of Freedom*, 20 mei 2019. Beschikbaar op https://www.bitsoffreedom.nl/2019/05/20/facebook-liegt-tegen-tweede-kamer-over-verkiezingsmanipulatie/

Autoriteit Persoonsgegevens (2020) 'Pas op met camera's met gezichtsherkenning', nieuwsbericht, 29 oktober 2020. Beschikbaar op: https://autoriteitpersoonsgegevens.nl/nl/nieuws/ap-pas-op-met-camera%E2%80%99s-met-gezichtsherkenning

AWTI (2020) *Krachtiger kiezen voor sleuteltechnologieën*, Den Haag: Adviesraad voor wetenschap, technologie en innovatie. Beschikbaar op: https://www.awti.nl/binaries/awti/documenten/adviezen/2020/01/30/

awti-advies-krachtiger-kiezen-voor-sleuteltechnologieen/Krachtiger+kiezen+voor+sleuteltechnologie%C3%ABn.pdf

Bakker, S. (2017) *From Luxury to Necessity: What the railways, electricity and automobile teach us about the it revolution*, Amsterdam: Boom Uitgeverij.

Bakker, S. en P. Korsten (2021) *Artificiële intelligentie als een general purpose technology: Strategische belangen van verantwoorde inzet in historisch perspectief*, WRR Working Paper nr. 41, Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid. Beschikbaar op: https://www.wrr.nl/publicaties/working-papers/2021/02/16/artificiele-intelligentie-als-een-general-purpose-technology

Barkhuysen, T. (2021) Handhaving van de AVG: de AP kan het niet alleen, *Nederlands Juristenblad*, 572(8).

Baruffaldi, S., B. van Beuzekom, H. Dernis, D. Harhoff, N. Rao, D. Rosenfield en M. Squicciarini (2020) *Identifying and measuring developments in artificial intelligence: Making the impossible possible*, OECD Science, Technology and Industry Working Papers 20(5), Parijs: OECD Publishing.

Bendett, S. (2019) *The Development of Artificial Intelligence in Russia*, blz. 168-198 in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Orde*r, Montgomery: Air University Press.

Benjamin, R. (2019) *The Race After Technology: Abolitionist Tools for the New Jim Code*, Cambridge: Polity Press.

Bennett Moses, L. (2007a) 'Why have theory of Law and Technological Change?', *Minnesota Journal of Law, Science & Technology*, 8: 589-606.

Bennett Moses, L. (2007b) 'Recuring Dilemmas: The Law`s Race to Keep Up with Technological Change', *University of Illinois Journal of Law, Technology and Policy*, vol. Fall: 239-285.

Bennett Moses, L. en N. Gollan (2015) *The Illusion of Newness: The Importance of History in Understanding the Law-Technology Interface*, UNSW Law Research Paper, 2015-71, Sydney: University of New South Wales.

Bernstein, G. (2006)' The Paradoxes of Technological Diffusion: Genetic Discrimination and Internet Privacy', *Connecticut Law Review*, 39(1), 243-297.

Bertolini, A. (2020) *Artificial Intelligence and Civil Liability. Onderzoek in opdracht van de Commissie Juridische Zaken van het Europese Parlement*, Brussel: Europees Parlement. Beschikbaar op: http://www.europarl.europa.eu/supporting-analyses

Bijker, W. (1995) *Of Bicycles, Bakelites, and Bulbs. Toward a Theory of Socio-technical Change*, Cambridge: MIT Press.

Bijlage bij de brief van staatssecretaris van Economische Zaken en Klimaat (2019, 17 mei) *Discussienotitie: toekomstbestendigheid mededingingsbeleid in relatie tot online platforms.* Beschikbaar op: https://www.rijksoverheid.nl/documenten/vergaderstukken/2019/05/20/bijlage-1-discussienotitie-toekomstbestendig-mededingsbeleid-in-relatie-tot-online-platforms

Bijlsma, M., B. Overvest en B. Straathof (2016) *Marktordening bij nieuwe ICT-toepassingen. Vroegtijdig ingrijpen nodig*, CPB Policy Brief 2016/09, Den Haag: Centraal Planbureau. Beschikbaar op: https://www.cpb.nl/sites/default/files/omnidownload/CPB-Policy-Brief-2016-09-Marktordening-bij-nieuwe-ICT-toepassingen.pdf

Bits of Freedom (z.d.) *Gezichtsherkenning*. Beschikbaar op https://www.bitsoffreedom.nl/dossiers/gezichtsherkenning/

Black, J. en A. Murray (2019) 'Regulating AI and machine learning: setting the regulatory agenda', *European Journal of Law and Technology*, 10(3). Beschikbaar op: https://ejlt.org/index.php/ejlt/article/download/722/980

Blankena, F. (2013) 'Identiteitskaart wordt mogelijk zonder vingerafdruk', *Binnenlandsbestuur Digitaal*, 9 september 2013. Beschikbaar op https://www.binnenlandsbestuur.nl/digitaal/nieuws/identiteitskaart-wordt-mogelijk-zonder.9097480.lynkx

Boden, M. (2018) *Artificial intelligence: A very short introduction*, Oxford: Oxford University Press.

Boom, W. van en F. Weber (2017) 'Collectief procederen – Ontwikkelingen in Nederland en Duitsland', *Weekblad voor Privaatrecht, Notariaat en Registratie*, WPNR 2017/7145: 291-299.

Bostrom, N. (2016) *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press.

Bousquet, A. (2009) *The scientific way of warfare: Order and Chaos on the Battlefields of Modernity*, Londen: HURST.

Bradford, A. (2020) *The Brussels effect: How the European Union rules the world*, Oxford: Oxford University Press.

Brand, S. (1995) 'We owe it all to the hippies', *Time Magazine*, 1 maart 1995. Beschikbaar op: http://content.time.com/time/subscriber/article/0,33009,982602,00.html

Bratton, B. (2016) *The stack: On software and sovereignty*, Cambridge: MIT Press.

Braun, C. en R. Stolk (2020) 'Procederen uit naam van het algemeen belang', *Montesquieu Instituut*, 4 maart 2020. Beschikbaar op: https://www.montesquieu-instituut.nl/id/vl6ck1v35y97/nieuws/procederen_uit_naam_van_het_algemeen

Bresnahan, T. en M. Trajtenberg (1995) 'General purpose technologies 'Engines of growth'?', *Journal of econometrics,* 65(1): 83-108.

Brooks, R. (2018) 'My dated predictions', blog, *Rodneybrooks*, 1 januari 2018. Beschikbaar op: https://rodneybrooks.com/my-dated-predictions/

Broussard, M. (2019) *Artificial Unintelligence: How Computers Misunderstand the World*, Cambridge: MIT Press.

Brown, R. (2021)' Property ownership and the legal personhood of artificial intelligence', *Information and Communications Technology Law*, 2: 208-234.

Brownsword, R. en M. Goodwin (2012) *Law and the Technologies of the Twenty-First Century*, Cambridge: Cambridge University Press.

Brundage, M., S. Avin, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, A. Dafoe, P. Scharre, T. Zeitzoff, B. Filar, H. Anderson, H. Roff, G. Allen, J. Steinhardt, C. Flynn, S. hÉigeartaigh, S. Beard, H. Belfield, S. Farquhar, C. Lyle, R. Crootof, O. Evens, M. Page, J. Bryson, R. Yampolskiy en D. Amodei (2018) *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation,* Oxford: Future of Humanity Institute.

Brynjolfsson, E. en A. McAfee (2014) *The Second Machine Age: Work, progress, and prosperity in a time of brilliant technologies*, New York: WW Norton en Company.

Brynjolfsson, E., D. Rock en C. Syverson (2019) 'Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics', blz. 23-57 in A. Agrawal, J. Gans en A. Goldfarb (2019) *The Economics of Artificial Intelligence: An Agenda*, Chicago: University of Chicago Press.

Buchem, M. van, H. Boosman, M. Bauer, I. Kant, S. Cammel en E. Steyerberg (2021) 'The digital scribe in clinical practice: a scoping review and research agenda', *NPJ Digital Medicin*, 4(57): 1-8.

Bughin, J., J. Seong, J. Manyika, M. Chui en R. Joshi (2018) *Notes From the AI Frontier: Modeling the Impact of AI on the World Economy*, McKinsey Global Institute. Beschikbaar op: https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.pdf?shouldIndex=false

Burrell, J. (2016) 'How the machine 'thinks': Understanding opacity in machine learning algorithms', *Big Data & Society* 3(1): 1-12.

Buruma, Y. (2020) 'International Law and Cyberspace. Issues of Sovereignty and the Common Good', blz. 69-111 in *International Law for a Digitalised World*, Den Haag: KNVIR/T.M.C. Asser Press.

Business Insider (2010) 'Mark Zuckerberg, Moving Fast and Breaking Things', *Businessinsider.nl.*, 15 oktober 2010. Beschikbaar op: https://www.businessinsider.com/mark-zuckerberg-2010-10?international=true&r=US&IR=T

Bygrave, L. (2021) 'The 'Strasbourg Effect' on data protection in light of the 'Brussels Effect': Logic, mechanics and prospects', *Computer Law & Security Review*, 40, 105460.

Campolo, A., M. Sanfilippo, M. Whittaker en K. Crawford (2017) *AI Now 2017 Report*, New York: AI Now Institute. Beschikbaar op: https://ainowinstitute.org/AI_Now_2017_Report.pdf

Cath, C., S. Wachter, B. Mittelstadt, M. Toledo en L. Floridi (2018) 'Artificial Intelligence and the 'Good society'. The US, EU, and UK approach', *Science and Engineering Ethics*, 24: 505-528.

CB Insights (2021, 24 juni) *Despite A Pandemic Slump, The AI Sector Remains Hot For Acquirers*, CB Insights Research Briefs. Beschikbaar op: https://www.cbinsights.com/research/artificial-acquisitions-trends-annual-deals/

CB Insights (2018, 26 april) *Rise of China's Big Tech in AI; What Baidu, Alibaba, and Tencent are Working on*, CB Insights Research Briefs. Beschikbaar op: https://www.cbinsights.com/research/china-baidu-alibaba-tencent-artificial-intelligence-dominance/

Centrale Raad van Beroep (2019) ECLI:NL:CRVB:2019:1737, uitspraak 15 mei 2019. Beschikbaar op: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:CRVB:2019:1737

CEG (2018) *Digitale dokters: Een ethische verkenning van medische expertsystemen*, Den Haag: Centrum voor Ethiek en Gezondheid. Beschikbaar op: https://www.ceg.nl/binaries/ceg/documenten/signalementen/2018/07/04/digitale-dokters---een-ethische-verkenning-van-medische-expertsystemen/webversie_CEG_Digitale_dokters_Een_ethische_verkenning_van_medische_expertsystemen.pdf

Chavannes, R., A. Strijbos en D. Verhulst (2021) 'Kroniek Recht en Technologie', *Nederlands Juristenblad* 2021, 16: 1350-1370. Beschikbaar op: https://blog.chavannes.net/2021/04/kroniek-technologie-en-recht-2021/

Chen, S. (2017) 'China to build giant facial recognition database to identify any citizen within seconds', *South China Morning Post*, 12 oktober 2017. Beschikbaar op: https://www.scmp.com/news/china/society/article/2115094/china-build-giant-facial-recognition-database-identify-any

Chiusi, F., S.Fischer, N. Kayser-Bril en M. Spielkamp (2020) *Automating Society Report 2020*, Berlijn: AlgorithmWatch. Beschikbaar op: https://www.ivir.nl/publicaties/download/Automating-Society-Report-

Choi, W., M. van Eck en H. Hukshorn (2021) *Hoe gemeenten besluiten over algoritmen en mensenrechten*, onderzoek in opdracht van het College voor de Rechten van de Mens, Den Haag: Hooghiemstra en partners.

Cihon, P. (2019) *Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development* Technical Report Oxford: Future of Humanity Institute.

Claus, S. (2017) 'Een algoritme dat aan je gezicht ziet of je homo of hetero bent', *Trouw*, 12 september 2017. Beschikbaar op: https://www.trouw.nl/nieuws/een-algoritme-dat-aan-je-gezicht-ziet-of-je-homo-of-hetero-bent~b7b99615/

Coeckelbergh, M. (2020) 'AI for climate: freedom, justice, and other ethical and political challenges', *AI and Ethics 1:* 67-72.

Coglianese, C. en D. Lehr (2019) 'Transparancy and Algorithmic Governance' *Administrative Law Review*, 71(1): 12-57.

College voor de Rechten van de Mens (2021) *Handreiking: (semi-)geautomatiseerde besluitvorming door de overheid*, Utrecht: College voor de Rechten van de Mens. Beschikbaar op: https://publicaties.mensenrechten.nl/file/002e83bf-47f8-44fd-846e-46aff40f8a56.pdf

Committee on the Judicary (2020) *Investigation of Competition in Digital Markets. Majority Staff Report and Recommendations*, Subcommitte on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary, Washington. Beschikbaar op: https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf

Crawford, K. (2021) *The Atlas of AI*, New Haven: Yale University Press.

Crawford, K., Dobbe, T. Dryer, G. Fried, B. Green, E. Kaziunas, A. Kak, V. Mathur, E. McElroy, A. Sánchez, D. Raji, J. Rankin, R. Richardson, J. Schultz, S. West en M. Whittaker (2019) *AI Now 2019 Report*, New York: AI Now Institute. Beschikbaar op: https://ainowinstitute.org/AI_Now_2019_Report.pdf

Creemers, R. (2019) 'The International and Foreign Policy Impact of China's Artificial Intelligence and Big-Data Strategies', blz. 129-135 in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, Montgomery: Air University Press.

Crémer, J., Y. De Montjoye en H. Schweitzer (2019) *Competition policy for the digital era*, Brussel: Europese Commissie. Beschikbaar op: http://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf

CSR (2018) *CSR Adviesbrief inzake opheffing 'Numeri Fixi'*, Den Haag: Cyber Security Raad. Beschikbaar op: https://www.cybersecurityraad.nl/binaries/cybersecurityraad/documenten/adviezen/2018/07/26/csr-adviesbrief-inzake-opheffing-%E2%80%98numeri-fixi/CSR+Adviesbrief+inzake+opheffing+%E2%80%98Numeri+Fixi%27.pdf

CSR (2020) *Naar structurele inzet van innovatieve toepassingen van nieuwe technologieën voor de cyberweerbaarheid van Nederland*, CSR-advies 2020, nr. 5, Den Haag: Cyber Security Raad. Beschikbaar op: https://www.cybersecurityraad.nl/adviezen/documenten/adviezen/2020/09/18/csr-advies-%E2%80%98naar-structurele-inzet-van-innovatieve-toepassingen-van-nieuwe-technologieen-voor-de-cyberweerbaarheid-van-nederland%E2%80%99---csr-advies-2020-nr.-5

CSR (2021) *Nederlandse Digitale Autonomie en Cybersecurity*, CSR-advies 2021, nr. 3, Den Haag: Cyber Security Raad. Beschikbaar op: https://www.cybersecurityraad.nl/documenten/adviezen/2021/05/14/csr-advies-nederlandse-digitale-autonomie-en-cybersecurity---csr-advies-2021-nr.-3

Danaher, J. (2016) 'The threat of algocracy: Reality, resistance and accomodation', *Philosophy & Technology*, 29(3): 245-268.

Danziger, S., J. Levav en L. Avnaim-Pesso (2011) 'Extraneous factors in judicial decisions', *Proceedings of the National Academy of Sciences*, 108(17): 6889-6892.

Das, D., R. de Jong en L. Kool, m.m.v. J. Gerritsen (2020) *Werken op waarde geschat – Grenzen aan digitale monitoring op de werkvloer door middel van data, algoritmen en AI*, Den Haag: Rathenau Instituut.

Data Science Center Tilburg (z.d.) *Data Science for Social Good (DS for Social Good)*. Beschikbaar op: https://www.tilburguniversity.edu/research/institutes-and-research-groups/data-science-center/dssg

Dauverge, P. (2020) *AI in the Wild - Sustainability in the Age of Artificial Intelligence,* Cambridge: MIT Press.

Davenport, C. (2019) *The Space Barons: Elon Musk, Jeff Bezos and the Quest to Colonize the Cosmos*, New York Public Affairs.

Davidson, D. en R. Delhaas (2020) 'Als de politiek in ieder oor een andere belofte fluistert', *Argos*, 22 april 2020. Beschikbaar op: https://www.vpro.nl/argos/lees/nieuws/2020/microtargeting-in-Nederland.html

De Conca, S. (2021) *The enchanted house. An analysis of the interaction of intelligent personal home assistants (IPHAS) with the private sphere and its legal protection*, dissertatie, Tilburg: Tilburg University.

De Poorter, J. en J. Goossens (2019) Effectieve rechtsbescherming bij algoritmische besluitvorming in het bestuursrecht. *Nederlands Juristenblad*, 44: 3303-3312.

De Ree, M. (2021) *Onderzoek naar eerlijke en uitlegbare algoritmen*, CBS.nl, 29 april 2021. Beschikbaar op: https://www.cbs.nl/nl-nl/corporate/2021/17/onderzoek-naar-eerlijke-en-uitlegbare-algoritmen

De Rijke, M. (2019) 'Investeer in kennisbasis AI of word een toeschouwer', *NRC*, 8 april 2019. Beschikbaar op: https://www.nrc.nl/nieuws/2019/04/08/investeer-in-kennisbasis-ai-of-wordt-een-toeschouwer-a3956136

Chen, Y., N. Casagrande, Y. Zhang en M. Brenner (2019) *Using WaveNet technology to reunite speech-impaired users with their original voices*, DeepMind.nl, 18 december. Beschikbaar op: https://deepmind.com/blog/article/Using-WaveNet-technology-to-reunite-speech-impaired-users-with-their-original-voices

DeepMind (2019) *Machine learning can boost the value of wind energy*, Deepmind.nl, 26 februari. Beschikbaar op https://deepmind.com/blog/article/machine-learning-can-boost-value-wind-energy

DenkWerk (2018) *Artificial Intelligence in Nederland: zelf aan het stuur.* Beschikbaar op: https://denkwerk.online/media/1029/artitificial_intelligence_in_nederland_juli_2018.pdf

Dennett, D. (2019) 'What Can We Do?', blz. 41-53 in J. Brockman (red.), *Possible Minds: Twenty-Five Ways of Looking at AI*, Londen: Penguin.

Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (2019) *Facebook-Auftritte von öffentlichen Stellen des Bundes*,

brief d.d. 20 mei 2019, 61924/2021. Beschikbaar op: https://www.bfdi.bund.de/SharedDocs/Downloads/DE/DokumenteBfDI/Rundschreiben/Allgemein/2021/Facebook-Auftritte-Bund.pdf?__blob=publicationFile&v=2

Dickson, B. (2020) 'Understanding the limits of CNNs, one of AI's greatest achievements', *TechTalks*, 2 maart 2020. Beschikbaar op https://bdtechtalks.com/2020/03/02/geoffrey-hinton-convnets-cnn-limits/

Dignum, V. (2019) *Responsible artificial intelligence: How to develop and use AI in a responsible way*, Cham: Springer Nature.

Dignum, V. (z.d.) *There is no AI – race and if there is, it's the wrong one to run*. Beschikbaar op: https://allai.nl/there-is-no-ai-race/

Dijck, G. van (2020) 'Algoritmische risicotaxatie van recidive. Over de Oxford Risk of Recidivism tool (OXREC), ongelijke behandeling en discriminatie in strafzaken', *Nederlands Juristenblad*, 25: 1784-1790.

Dijck, J. van (2020) 'Seeing the forest for the trees: Visualizing platformization and its governance', *New Media & Society*, 00(0): 1-19.

Dijksterhuis, E. (2006 [1950]) *De mechanisering van het wereldbeeld*, Amsterdam: Amsterdam University Press.

Ding, J. (2018) *Deciphering China's AI dream*, Future of Humanity Institute Technical Report, Oxford: University of Oxford.

Ding, J. (2019) 'The Interests behind China's Artificial Intelligence Dream', blz. 43-47 in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, Montgomery: Air University Press.

Diogo, M. en D. van Laak, (2016) *Europeans Globalizing: Mapping, Exploiting, Exchanging*, Londen: Palgrave Macmillan.

Domingos, P. (2017) *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Londen: Penguin Random House.

Domini, A. en Chicot, J. (2018) *Case Study Report: From Concorde to Airbus, report for the European Commission*, Brussel: Europese Commissie. Beschikbaar op: http://publications.europa.eu/resource/cellar/4940e0c9-2359-11e8-ac73-01aa75ed71a1.0001.01/DOC_1

Dommering, E. (red.) (2000) *Informatierecht: fundamentele rechten voor de informatiesamenleving*, Amsterdam: Otto Cramwinckel.

Dreyfus, H. en S. Dreyfus (1986) *Mind over Machine*, New York: The Free Press.

Drezner, D. (2019) 'Economic Statecraft in the Age of Trump', *The Washington Quarterly*, 42(3): 7-24.

Dutton, T. (2018) 'An Overview of National AI Strategies', *Medium*, 28 juni 2018. Beschikbaar op: https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

Eck, M. van, S. Zouridis en M. Bovens (2018) 'Algoritmische rechtstoepassing in de democratische rechtsstaat', *Nederlands Juristenblad*, 40: 3008-3017.

Edgerton, D. (2008) *The Shock of the Old: Technology and global history since 1900*, Londen: Profile books.

El-Dardiry, R., B. Overvest, M. Dinkova en R. Albers (2021) *Brave new data. Databeleid in een imperfecte wereld*, CPB Policy Brief, mei 2021, Den Haag: Centraal Planbureau. Beschikbaar op: https://www.cpb.nl/brave-new-data-databeleid-in-een-imperfecte-wereld

Ettekoven, B.J. van en C. Prins (2018) 'Data analysis, artificial intelligence and the judiciary system', blz. 425-447 in V. Mak, E. Tjong Tjin Tai en A. Berlee (reds.) *Research Handbook in Data Science and Law*, Cheltenham: Edward Elgar Publishing.

Eubanks, V. (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York: St. Martin's Press.

European Center for Not-for-Profit Law (2020) *Being aware: incorporating civil society into national strategies on artificial intelligence. Country papers on participatory processes in drafting national AI policies in the Czech Republic, the Netherlands, Australia and Canada*, Brussel: ECNPL. Beschikbaar op: https://ecnl.org/publications/being-ai-ware-incorporating-civil-society-national-strategies-artificial-intelligence

European Data Protection Board en European Data Protection Supervisor (2021) *Joint opinion 5/2021 on the proposal for a proposal for the regulation of the European parliament and of the council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act), 28 juni 2021*, Brussel: EDPB, EDPS. Beschikbaar op: https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf

European Political Strategy Centre (2018) *The Age of Artificial Intelligence: Towards a European Strategy for Human-Centric Machines*, EPSC Strategic Notes, Brussel: EPSC. Beschikbaar op: https://ec.europa.eu/jrc/communities/en/community/digitranscope/document/age-artificial-intelligence-towards-european-strategy-human-centric

European Union Agency for Fundamental Rights (2020) *Getting the future right. Artificial Intelligence and fundamental rights*, Luxembourg: Publications Office of the European Union. Beschikbaar op: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf

Europees Parlement (2020) *European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL))*. Beschikbaar op: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html

Europese Commissie (2018) *Member States and Commission to work together to boost artificial intelligence "made in Europe"*, persbericht, 7 december2018. Beschikbaar op https://ec.europa.eu/commission/presscorner/detail/en/IP_18_6689

Europese Commissie (2018a) *Annex to the Coordinated Plan on the Development and Use of Artificial Intelligence Made in Europe*, Brussel: Europese Commissie. Beschikbaar op: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56017

Europese Commissie (2020) *A European Strategy for Data*, COM(2020) 66 final. Beschikbaar op: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX:52020DC0066

Europese Commissie (2021) *2030 Digital Compass: the European way for the Digital Decade*. Beschikbaar op: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52021DC0118

Europese Commissie (2021a [2018]) *EU Coordinated Action Plan on AI 2021 Review*, COM(2021) 205 final. Beschikbaar op: https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review

Europese Commissie (2021b) *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM(2021) 206 final. Beschikbaar op https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

Europese Commissie (z.d. a) *Destination Earth*. Beschikbaar op: https://digital-strategy.ec.europa.eu/en/policies/destination-earth

Europese Commissie (z.d. b) *Public Consultation: White Paper on Artificial Intelligence – a European Approach*. Beschikbaar op: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12270-White-Paper-on-Artificial-Intelligence-a-European-Approach/public-consultation_en

Feldstein, S. (2019) *The Global Expansion of AI Surveillance*, Washington: Carnegie Endowment for International Peace.

Fiebig, T., S. Gürses, C. Gañán, E. Kotkamp, F. Kuipers, M. Lindorfer, M. Prisse en T. Sari (2021) *Heads in the Clouds: Measuring the Implications of Universities Migrating to Public Clouds*. Beschikbaar op: https://arxiv.org/abs/2104.09462

Field, A. (2008) *Does economic history need GPTS?* Beschikbaar op: http://ssrn.com/abstract=1275023

Fierens, M., E. van Gool, en De Bruyne, J. (2021) 'De regulering van artificiële intelligentie (deel 1) – Een algemene stand van zaken en een analyse van enkele vraagstukken inzake consumentenbescherming', *Rechtskundig Weekblad*, 84(25): 962-980.

Fight for the Future (z.d.) Interactieve kaart. Beschikbaar op: https://www.banfacialrecognition.com/map/

FLIR UGS (z.d.) *Combat-Proven Robots*. Beschikbaar op: https://www.flir.com/uis/ugs/

Floridi, L. (2014) *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford: Oxford University Press.

Floridi, L. (2021) 'The European Legislation on AI: A Brief Analysis of its Philosophical Approach', *Philosophy and Technology*, 34: 215-222. Beschikbaar op: https://link.springer.com/article/10.1007/s13347-021-00460-9

Floridi, L. en J. Cowls (2019) 'A Unified Framework of Five Principles for AI in Society', *Harvard Data Science Review*, 1(1). Beschikbaar op: https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., M. Taddeo en M. Turilli (2009) 'Turing's imitation game: still an impossible challenge for all machines and some judges–an evaluation of the 2008 Loebner contest', *Minds and Machines*, 19(1), 145-150.

Floridi, L., J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke en E. Vayena (2018) 'AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations', *Minds and Machines*, 28: 689-707.

Ford, M. (2018) *Architects of Intelligence*, Birmingham: Packt Publishing.

France (2018) *AI for Humanity: French Strategy for Artificial Intelligence,* President of the French Republic. Beschikbaar op: https://www.aiforhumanity.fr/en/

Franken, H. (1993) 'Kanttekeningen bij het automatiseren van beschikkingen', blz. 7-50 in H. Franken, I.TH.M. Snellen, J. Smit en A.W. Venstra *Beschikken en automatiseren*, pre-advies Vereniging voor Administratief Recht, Alphen a/d Rijn: Samsom H.D. Tjeenk Willink.

Freeman, S. (2001) 'Illiberal Libertarians', *Philosophy & Public Affairs*, 30(2): 105-151.

Freeman, C. en F. Louçã (2001) *As Time Goes By: From the Industrial Revolutions to the Information Revolution*, Oxford: Oxford University Press.

Freeman, K., J. Dinnes, N. Chuchu, Y. Takwoingi, S.E. Bayliss, R. Matin, A. Jain, F. Walter, H. Williams en J. Deeks (2020) 'Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies', *BMJ*, 368(127).

Frenkel, S. (2018) 'Microsoft employees protest work with ICE, as Tech industries mobilizes over immigration', *The New York Times*, 19 juni 2018. Beschikbaar op: https://www.nytimes.com/2018/06/19/technology/tech-companies-immigration-border.html#:~:text=%E2%80%9CWe%20believe%20that%20Microsoft%20must,the%20chief%20executive%2C%20Satya%20Nadella.&text=The%20policy%20has%20resulted%20in,parents%2C%20raising%20a%20bipartisan%20outcry.

Frenken, K. en L. Fuenfenschilling (2020) 'The Rise of Online Platforms and the Triumph of the Corporation', *Sociologica*, 14(3): 101-113.

Fridman, L. (2019) *Elon Musk: What's Outside the Simulation?*, AI Podcast Clips, 16 augustus 2019. Beschikbaar op: www.youtube.com/watch?v=YIVf3P3zq7g

Fukuyama, F. (1992) *The End of History and the Last Man*, Toronto: Simon & Schuster.

Fukuyama, F. (2021) 'Making the Internet Safe for Democracy', *Journal of Democracy*, 32(2): 37-44.

Fukuyama, F., B. Richman, A. Goel, M. Schaake, R. Katz en D. Melamed (2021) *Report of the working group on platform scale*, Stanford: Stanford University. Beschikbaar op: https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/platform_scale_whitepaper_-cpc-pacs.pdf

Fussell, S. (2019) 'Why Hong Kongers Are Toppling Lampposts', *The Atlantic*, 30 augustus 2019. Beschikbaar op: https://www.theatlantic.com/technology/archive/2019/08/why-hong-kong-protesters-are-cutting-down-lampposts/597145/

Future of Life Institute (z.d. a) *National and International AI Strategies*, Future of Life Institute. Beschikbaar op: https://futureoflife.org/national-international-ai-strategies/?cn-reloaded=1&cn-reloaded=1

Future of Life Institute (z.d. b) *An open letter: Research proprities for robust and beneficial artificial intelligence*, Future of Life Institute. Beschikbaar op: https://futureoflife.org/ai-open-letter/

Future of Life Institute (z.d. c) *Asilomar AI principles*, Future of Life Institute. Beschikbaar op: https://futureoflife.org/ai-principles/

Gerbrandy, A. en B. Custers (2018) 'Algoritmische besluitvorming en het kartelverbod', *Markt en Mededinging*, 3: 101-109. Beschikbaar op: https://www.bjutijdschriften.nl/tijdschrift/marktenmededinging/2018/3/MenM_1387-6236_2018_021_003_002

Gerechtshof Amsterdam (2021a) ECLI:NL:GHAMS:2021:1560 – Gerechtshof Amsterdam, 01-06-2021 / 200.280.852/01. Beschikbaar op: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:GHAMS:2021:1560

Gerechtshof Amsterdam (2021b) ECLI:NL:GHAMS:2021:392. Beschikbaar op: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:GHAMS:2021:392

Gerritsen, J., J. Hamer, L. Kool en P. Verhoef (2020) 'Beter beschermd tegen biometrie', *Beleid en Maatschappij*, 47(4): 451-466.

Gezondheidsraad (2006) *Betekenis van nanotechnologieën voor de gezondheid*, Den Haag: Gezondheidsraad.

Giesea, J. (2016) 'It's time to embrace memetic warfare', *Defence Strategic Communications*, 1: 67-75.

Giesen, I. (2007) *Alternatieve regelgeving en privaatrecht*, Deventer: Kluwer.

Giesen, I. (2018) '(Zelf)Regulering van en in het privaatrecht; op zoek naar een 'ReL'?, *Nederlands Tijdschrift voor Burgerlijk Recht*, 5: 135.

Ginkel, J. van en P. Strijp (2020) 'Van beleidscyclus naar datacyclus', *iBestuur*. Beschikbaar op: https://ibestuur.nl/podium/van-beleids-naar-datacyclus

Goode, L. (2018) 'Google CEO says AI will be more important to humanity than electricity or fire', *The Verge*, 19 januari 2018. Beschikbaar op: https://www.theverge.com/2018/1/19/16911354/google-ceo-sundar-pichai-ai-artificial-intelligence-fire-electricity-jobs-cancer

Gool, E. van, J. de Bruyne en M. Fierens (2021) 'De regulering van artificiële intelligentie (deel 2). Een analyse van buitencontractuele aansprakelijkheid', *Rechtskundig Weekblad*, 84(26): 1003-1024.

Goossens, J., E. Hirsch Ballin en E. van Vugt (2021) 'Algoritmische beslisregels vanuit constitutioneel oogpunt. Tweedeling tussen algemene regels en concrete toepassing onder druk', *Tijdschrift voor constitutioneel recht*, 12(1): 4-19.

Gordon, R. (2016) *The rise and fall of American growth: The U.S. standard of living since the Civil War*, Princeton: Princeton University Press.

GPT-3 (2020) A robot wrote this entire article. Are you scared yet, human?', *The Guardian*, 8 september 2020. Beschikbaar op https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

Graef, I. en J. Prüfer (2021) 'Governance and Data Sharing: a Law en Economics Proposal', *Research Policy*, 50(9): 104330.

Greene, D., A. Hoffman en L. Stark (2019) 'Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning', blz. 2122-2131 in *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Greenfield, A. (2017) *Radical technologies: The design of everyday life*, New York: Verso Books.

Gross, A., M.Murgia en Y. Yang (2019) 'Chinese tech groups shaping UN facial recognition standards', *Financial Times*, 1 december 2019. Beschikbaar op: https://www.ft.com/content/c3555a3c-0d3e-11ea-b2d6-9bf4d1957a67

Hage, J. (2017) 'Theoretical foundations for the responsibility of autonomous agents', *Artificial Intelligence and Law,* 25(3): 255-271.

Hage, J. en B. Verheij (1999) 'Rechtsinformatica: de stand van zaken in de wetenschap', blz. 65-92 in A. Oskamp en A. Lodder (reds.), *Informatietechnologie voor juristen. Handboek voor de jurist in de 21e eeuw*, Deventer: Kluwer.

Hagedoorn P. (2021) *The Digital Challenge for Europe*. Peter Hagedoorn.

Hagendorff, T. (2020) 'The Ethics of AI Ethics: An Evaluation of Guidelines', *Minds and Machines* 30: 99-120.

Hall, P. en D. Soskice (2001) *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*, Oxford: Oxford University Press.

Halpern, S. (2019) 'The terrifying potential of the 5G network', *The New Yorker*, 26 april 2019. Beschikbaar op: www.newyorker.com/news/annals-of-communications/the-terrifying-potential-of-the-5g-network

Harari, Y.N. (2019) 'Who will win the race for AI?', *Foreign Policy Magazine*, Winter 2019. Beschikbaar op: https://foreignpolicy.com/gt-essay/who-will-win-the-race-for-ai-united-states-china-data/

Häußermann, J. en C. Lütge (2021) 'Community-in-the-loop: towards pluralistic value creation in AI, or—why AI needs business ethics', *AI and Ethics*: 1-22.

Hayek, F. (1994) *The Road to Serfdom*, Chicago: University of Chicago Press.

Heest, F. (2020) 'Nederland moet meer doen om talentvlucht te voorkomen bij kunstmatige intelligentie', *ScienceGuide*, 3 maart 2020. Beschikbaar op: https://www.scienceguide.nl/2020/03/een-belgische-postdoc-verdient-meer-

Helberger, N., J. Pierson en T. Poell (2018) 'Governing Online Platforms: From Contested to Cooperative Responsibility', *The Information Society* 34(1): 1-14.

Helbing, D., B. Frey, G. Gigenrenzer, E. Hafen, M. Hagner, Y. Hofstetter, J. van den Hoven, R. Zicari en A. Zwitter (2019) 'Will Democracy Survive Big Data and Artificial Intelligence?', blz. 73-98 in D. Helbing (red.), *Towards Digital Enlightenment*, Cham: Springer.

Helwig, P. (2020) 'Rekenen en rekenschap. Algoritmes en de Archiefwet', *Tijdschrift voor Toezicht*, 1: 54-59.

Hern, A. (2019) 'Tim Berners-Lee on 30 years of the world wide web: 'We can get the web we want'', *The Guardian*, 12 maart 2019. Beschikbaar op: https://www.theguardian.com/technology/2019/mar/12/tim-berners-lee-on-30-years-of-the-web-if-we-dream-a-little-we-can-get-the-web-we-want

Hern, A. (2019) 'Cambridge Analytica did work for Leave.EU, emails confirm', *The Guardian*, 30 juli 2019. Beschikbaar op: https://www.theguardian.com/uk-news/2019/jul/30/cambridge-analytica-did-work-for-leave-eu-emails-confirm

Heukelom-Verhage, S. van (2020) 'Maatwerk bieden in een gedigitaliseerde en datagedreven samenleving #HoeDan?', in L. van den Berge, M. Vermaat, M. Lurks, N. van Renssen en S. van Heukelom-Verhage (reds.) *Maatwerk in het bestuursrecht*, Den Haag: Boom.

Hicks, M. (2018)' Why tech's gender problem is nothing new', *The Guardian*, 12 oktober 2018. Beschikbaar op: https://www.theguardian.com/technology/2018/oct/11/tech-gender-problem-amazon-facebook-bias-women

High-Level Expert Group on Artificial Intelligence (2019) *A definition of AI: Main capabilities and scientific disciplines*, Brussel: Europese Commissie. Beschikbaar op: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341

High-Level Expert Group on Artificial Intelligence (2019a) *Ethics guidelines for trustworthy AI*, Brussel: Europese Commissie. Beschikbaar op: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

High-Level Expert Group on Artificial Intelligence (2019b) *Policy and Investment Recommendations for Trustworthy AI*, Brussel: Europese Commissie. Beschikbaar op: https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence

High-Level Expert Group on Artificial Intelligence (2020) *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*, Brussel: Europese Commissie. Beschikbaar op: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

HiiL (z.d.) *Supporting Justice Innovations*, Den Haag: The Hague Institute for Innovation of Law. Beschikbaar op: https://www.hiil.org/what-we-do/the-justice-accelerator/

Hildebrandt, M. (2018) 'Law As Computation in the Era of Artificial Legal Intelligence. Speaking Law to the Power of Statistics', *The University of Toronto Law Journal*, 68(5): 12-35.

Hildebrandt, M. en S. Gutwirth (reds.) (2008) *Profiling the European Citizens*, Berlin: Springer.

Hillman, J. (2019) *Infrastructure and Influence: The Strategic Stake of Foreign Projects*, Washington: Center for Strategic and International Studies.

Hirsch Ballin, E. (2021) *Mensenrechten als ijkpunten van artificiële intelligentie*, WRR Working Paper nr. 46, Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid.

Hirsch Ballin, E., A. Jaspers, A. Knottnerus en H. Vinke (2021) *De Toekomst van de sociale zekerheid: de menselijke maat in een solidaire samenleving*, Den Haag: Boom.

Hoffman, S. (2019) 'Managing the State: Social Credit, Surveillance, and the Chinese Communist Party's Plan for China', blz. 48-57 in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, Montgomery: Air University Press.

Hoge Raad (1921) NJ 1921, 564 (Elektriciteitsarrest) ECLI:NL:HR:1921:186, uitspraak 23 mei 1921. Beschikbaar op: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:HR:1921:186

Hoge Raad (2018) ECLI:NL:HR:2018:1316, uitspraak 17 augustus 2018. Beschikbaar op: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:HR:2018:1316

HoloniQ (2020) *The 2020 AI Strategy Landscape: 50 National Artificial Intelligence Strategies Shaping the Future of Humanity*, 9 april 2020. Beschikbaar op: https://www.holoniq.com/notes/50-national-ai-strategies-the-2020-ai-strategy-landscape/

Horowitz, M., G. Allen, E. Kania en P.Scharre (2018) *Strategic competition in an era of artificial intelligence. Center for a New American Security*, Washington: CNAS. Beschikbaar op: https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/

CNAS-Strategic-Competition-in-an-Era-of-AI-July-2018_v2.pdf?mtime=20180716122000enfocal=none
Houwerzijl, M. (2018) 'Juridische vraagstukken rond arbeid in de klusseneconomie', *Beleid en Maatschappij*, 45(2): 208-216.
Hoven, J. van den (2013) 'Value sensitive design and responsible innovation', blz. 85-107 in R. Owen, J. Bessant en M. Heintz (reds.) *Responsible innovation: Managing the responsible emergence of science and innovation in society*, Chichester: John Wiley & Sons.
Hueck, H. (2018) 'Kopstuk van kunstmatige intelligentie vertrekt naar Zweden', *Het Financieele Dagblad*, 2 september 2018. Beschikbaar op: https://fd.nl/ondernemen/1267790/kopstuk-van-kunstmatige-intelligentie-vertrekt-naar-zweden
Hueck, H. en J.F. van Wijnen (2019) 'Kabinet vaag over extra budget voor kunstmatige intelligentie', *Het Financieele Dagblad*, 8 oktober 2019. Beschikbaar op: https://fd.nl/economie-politiek/1319532/kabinet-wil-meer-geld-uittrekken-voor-kunstmatige-intelligentie
Huntington, S. (1991) *The Third Wave: Democratization in the Late 20th Century*, Oklahoma: University of Oklahoma press.
Huys, I., G. van Overwalle en G. Matthijs (2011) 'Gene and genetic diagnostic method patent claims: a comparison under current European and US patent law', *European Journal of Human Genetics,* 19(10): 1104-1107.
iBestuur (2021) 'Overleg met informateur Mariëtte Hamer over digitalisering', *iBestuur*, 31 mei 2021. Beschikbaar op: https://ibestuur.nl/nieuws/overleg-met-informateur-mariette-hamer-over-digitalisering
Ihde, D. (2010) *Embodied technics*, Kopenhagen: Automatic Press/VIP.
Ipsos (2019) *Ipsos Global Poll for the World Economic Forum Shows Widespread Concern about Artificial Intelligence*. Beschikbaar op: www.ipsos.com/sites/default/files/ct/news/documents/2019-07/wef-ai-ipsos-press-release-jul-1-2019_0.pdf
Jeffries, A. (2014) 'This anarchist collective is demanding $3 billion from Google.The Counterforce, Kevin Rose, and the fire beneath San Francisco's growing class gap', *The Verge*, 15 april 2014. Beschikbaar op: https://www.theverge.com/2014/4/15/5614652/deny-the-machine
Jobin, A., M. Ienca en E. Vayena (2019) 'The global landscape of AI ethics guidelines', *Nature Machine Intelligence*, 1(9): 389-399.
Johnson, C. (1982) *MITI and the Japanese miracle: the growth of industrial policy, 1925-1975*, Stanford: Stanford University Press.
Joler, V. en K. Crawford (2018) *Anatomy of an AI-system: An anatomical case study of the Amazon echo as a artificial intelligence system made of human labor*. Beschikbaar op: https://anatomyof.ai/img/ai-anatomy-map.pdf
Juma, C. (2016) *Innovation and its enemies: Why people resist new technologies*, Oxford: Oxford University Press.

Just, N. en M. Latzer (2017) 'Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet', *Media, Culture en Society* 39(2): 238-258.

Kaiser, W. en J. Schot (2014) *Writing the Rules for Europe: Experts, Cartels and International Organizations*, Londen: Palgrave Macmillan.

Kamerstukken II 2013/2014 26 643, nr. 298 (2013) *Vrijheid en veiligheid in de digitale samenleving. Een agenda voor de toekomst*, Kamerbrief, 13 december 2013. Beschikbaar op: https://zoek.officielebekendmakingen.nl/kst-26643-298.pdf

Kamerstukken II 2015/2016 34300-X, nr. 88 (2016) *Reactie op het advies 'Autonome wapensystemen: de noodzaak van betekenisvolle menselijke controle' van de Adviesraad Internationale Vraagstukken (AIV) en de Commissie van Advies inzake Volkenrechtelijke Vraagstukken (CAVV)*, Kamerbrief, 4 maart 2016. Beschikbaar op: https://www.tweedekamer.nl/downloads/document?id=498139ae-0353-4c0e-9678-68f3eb2da7e9&title=Reactie%20op%20het%20advies%20%E2%80%98Autonome%20wapensystemen%3A%20de%20noodzaak%20van%20betekenisvolle%20menselijke%20controle%E2%80%99%20van%20de%20Adviesraad%20Internationale%20Vraagstukken%20%28AIV%29%20en%20de%20Commissie%20van%20Advies%20inzake%20Volkenrechtelijke%20Vraagstukken%20%28CAVV%29.pdf

Kamerstukken II 2017/2018 32 761, nr. 117 (2018) *Verwerking en bescherming persoonsgegevens*, Motie, 6 juni Beschikbaar op: https://zoek.officielebekendmakingen.nl/kst-32761-117.pdf

Kamerstukken II 2018/2019, 26643, nr. 570 (2018) *Brief van de minister voor rechtsbescherming*, Kamerbrief, 9 oktober 2018. Beschikbaar op: https://zoek.officielebekendmakingen.nl/kst-26643-570.pdf

Kamerstukken II 2018/2019, 35 134, nr. 2. (20190) *Initiatiefnota van het lid Verhoeven over mededinging in de digitale economie*, Initiatiefnota, 5 februari 2019. Beschikbaar op: https://zoek.officielebekendmakingen.nl/kst-35134-2.pdf

Kamerstukken II 2018/2019 32 761, nr. 138 (2019) *Verwerking en bescherming persoonsgegevens*, Motie, 25 juni. Beschikbaar op: https://www.tweedekamer.nl/downloads/document?id=dd94a468-cf9d-4b5d-83a8-f608b99c2fc9&title=Motie%20van%20het%20lid%20Buitenweg%20over%20het%20voorkomen%20van%20indirecte%20discriminatie%20door%20besluitvorming%20via%20algoritmen.pdf

Kamerstukken II 2018/2019 32 761, nr. 152 (2019) *Waarborgen en kaders bij gebruik gezichtsherkenningstechnologie*, Brief regering, 20 november 2019. Beschikbaar op: https://www.tweedekamer.nl/downloads/document?id=1352f4b1-696c-499a-ab6d-67c2464ad6ff&title=Waarborgen%20en%20kaders%20bij%20gebruik%20gezichtsherkenningstechnologie.pdf

Kamerstukken II 2018/2019 33009, nr. 70 (2019) *Aanpak sleuteltechnologieën*, Bijlage bij Kamerbrief, 26 april 2019. Beschikbaar op: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/publicaties/2019/04/26/aanpak-sleuteltechnologieen/Bijlage+2+Kamerbrief+Missiegedreven+Topsectoren-+en+Innovatiebeleid.pdf

Kamerstukken II 2018/2019 2018Z22902 (2018) *Mondelinge vragen van het lid Alkaya (SP) aan de Staatssecretaris van Binnenlandse Zaken en Koninkrijksrelaties over het gebrek aan controle op het gebruik van algoritmen bij de overheid*, Mondelinge vragen, 4 december 2018. Beschikbaar op: https://www.tweedekamer.nl/downloads/document?id=8bcf76b1-416f-4583-832a-71e72492ee75&title=Het%20gebrek%20aan%20controle%20op%20het%20gebruik%20van%20algoritmen%20bij%20de%20overheid%20%28Binnenlandsbestuur.nl%2C%2029%20november%202018%29%20.pdf

Kamerstukken II 2018/2019 35 212, nr. 2 (2019) *Initiatiefnota van het lid Middendorp: menselijke grip op algoritmen*, Initiatiefnota, 29 mei 2019. Beschikbaar op: https://www.tweedekamer.nl/downloads/document?id=06cee835-6c90-4bdc-b067-fd28dfe2da31&title=Initiatiefnota%20.pdf

Kamerstukken II 2019/2020, 26643 nr. 641 (2019) *Brief van de minister voor Rechtsbescherming*, Kamerbrief, 8 oktober 2019. Beschikbaar op: https://zoek.officielebekendmakingen.nl/kst-26643-641.pdf

Kamerstukken II 2019/2020 26 643, nr. 641 (2019) *Waarborgen tegen risico's van data-analyses door de overheid*, Kamerbrief, 8 oktober 2019. Beschikbaar op: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2019/10/08/tk-waarborgen-tegen-risico-s-van-data-analyses-door-de-overheid/tk-waarborgen-tegen-risico-s-van-data-analyses-door-de-overheid.pdf

Kamerstukken II 2019/2020 26 643, nr. 642 (2019) *AI, publieke waarden en mensenrechten*, Brief regering, 8 oktober 2019. Beschikbaar op: https://www.tweedekamer.nl/downloads/document?id=1a5131a6-0f2e-4b5e-917f-8f3e2cf0a144&title=AI%2C%20publieke%20waarden%20en%20mensenrechten.pdf

Kamerstukken II 2019/2020 26 643, nr. 652 (2019) *Artificiële Intelligentie bij de politie*, Kamerbrief, 3 december 2019. Beschikbaar op: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2019/12/03/tk-artificiele-intelligentie-bij-de-politie/tk-artificiele-intelligentie-bij-de-politie.pdf

Kamerstukken II 2019/2020 26 643, nr. 681 (2020) *Verslag van een Algemeen Overleg*, Verslag, 12 maart 2020. Beschikbaar op: https://zoek.officielebekendmakingen.nl/kst-26643-681.pdf

Kamerstukken II 2019/2020 30950, nr. 206 (2020) *Rassendiscriminatie*, Motie, 1 juli 2020. Beschikbaar op: https://zoek.officielebekendmakingen.nl/kst-30950-206.html

Kamerstukken II 2019/2020 26 643 en 32 761, nr. 652 (2019) *Beantwoording schriftelijke vragen AI bij de politie*, Kamerbrief, 3 december 2019. Beschikbaar op: https://zoek.officielebekendmakingen.nl/kst-26643-652.pdf

Kamerstukken II 2019/2020, 26643, nr. 672 (2020) *Brief van de ministers van Binnenlandse Zaken en Koninkrijksrelaties en voor Rechtsbescherming*, Kamerbrief, 13 maart 2020. Beschikbaar op: https://zoek.officielebekendmakingen.nl/kst-26643-672.html

Kamerstukken II 2020/2021, 26 643, nr. 765 (2021) *Voortgangsbrief AI en algoritmen*, Kamerbrief, 10 juni 2021. Beschikbaar op: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2021/06/10/kamerbrief-voortgang-algoritmen-en-artificiele-intelligentie/kamerbrief-over-algoritmen-en-artificiele-intelligentie-ai.pdf

Kamerstukken II 2020/2021, 28362, nr. 44 (2021) *Motie van de leden Klaver en Ploumen: Reikwijdte van artikel 68 Grondwet*, Motie, 29 april 2021. Beschikbaar op: https://zoek.officielebekendmakingen.nl/kst-28362-44.pdf

Kamerstukken II 2020/2021, 26643, nr. 779 (2021) *Informatie- en communicatietechnologie (ICT); Brief regering; Nieuwe I-strategie Rijk 2021-2025*, 7 september 2021. Beschikbaar op: https://zoek.officielebekendmakingen.nl/kst-26643-779.html

Kasparov, G. (2018) *Deep Thinking. Where Machine Intelligence Ends and Human Creativity Begins*, Londen: John Murray Press.

Kayser-Bril, N. (2019) 'At least 11 police forces use face recognition in the EU, AlgorithmWatch reveals', *AlgorithmWatch*, 11 december 2019. Beschikbaar op: https://algorithmwatch.org/en/face-recognition-police-europe/

Keane, J. (2009) *Life and death of democracy*, Toronto: Simon & Schuster.

Keane, J. (2011) 'Monitory democracy', blz. 212-235 in S. Alonso (red.) *The future of representative democracy*, Cambridge: Cambridge University Press.

Kelly, K. (2017) *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future*, Londen: Penguin.

Kerr, J. (2019) 'The Russian Model of Digital Control and Its Significance', blz. 62-74 in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, Montgomery: Air University Press.

Keymolen, E., M. Noorman, B. van der Sloot, B.-J. Koops, C. Cuijpers en B. Zhao (2020) *Op het eerste gezicht: Een verkenning van gezichtsherkenning en privacyrisico's in horizontale relaties*, Den Haag: Wetenschappelijk Onderzoek- en Documentatiecentrum. Beschikbaar op: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/

rapporten/2020/04/20/tk-bijlage-wodc-rapport-op-het-eerste-gezicht/tk-bijlage-

Keynes, J. M. (1930) 'Economic possibilities for our grandchildren', blz. 358-373 in J.M. Keynes (1932) *Essays in Persuasion*, New York: Harcourt Brace.

Kleinberg, J. (2018) 'Inherent Trade-Offs in Algorithmic Fairness', blz. 40 in *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '18)*, New York: ACM.

Klincewicz, M. en F. Lily (2020) *Consequences of unexplainable machine learning for the notions of a trusted doctor and patient autonomy*, paper presented at the 32nd International Conference on Legal Knowledge and Information Systems, Madrid, Spain, 2020.

Kohlen, J., M. van de Sande en M. Cox (2021) ''Rebooting' het mededingingsrecht – ook het mededingingsrecht ontsnapt niet aan de digitale transitie', *Markt en Mededinging*, 1: 6-14.

Kool, L., J. Timmer, L. Royakker en R. van Est (2017) *Opwaarderen: Borgen van publieke belangen in de digitale samenleving*, Den Haag: Rathenau Instituut.

Koops, B. (2006) 'Should ICT Regulation be Technology-Neutral', blz. 77-108 in B.J. Koops, C. Prins, M. Schellekens en M. Lips (reds.), *Starting points for ICT regulation. Deconstructing prevalent policy one-liners*, Den Haag: T.M.C. Asser Press.

Koops, B.J., M. Lips, J. Nouwt, C. Prins en M. Schellekens (2006) 'Should selfregulation be the starting point?', blz. 109-149 in B.J. Koops, C. Prins, M. Schellekens en M. Lips (reds.) (2006), *Starting points for ICT regulation. Deconstructing prevalent policy one-liners*, Den Haag: T.M.C. Asser Press.

Kop, M. (2020) 'AI and Intellectual Property: Towards an Articulated Public Domain', *Texas Intellectual Property Law Journal*, 28(1): 297-341.

Kosta, E. (2020) 'Algorithmic state surveillance: Challenging the notion of agency in human rights', *Regulation & Governance*. http://dx.doi.org/10.1111/rego.12331

Krupiy, T. (2020) 'A Vulnerability Analysis: Theorising the Impact of Artificial Intelligence Decision-Making Processes on Individuals, Society and Human Diversity from a Social Justice Perspective', *Computer Law & Security Review*, 38: 1-25.

Kuijpers, K., T. Muntz en T. Staal (2018) 'Privacy? Achterhaald', *De Groene Amsterdammer*, 31 oktober 2018. Beschikbaar op: https://www.groene.nl/artikel/privacy-achterhaald

Kulk, S. (2020) 'Platformaansprakelijkheid – van 'notice and takedown' naar algoritmisch toezicht', *Nederlands tijdschrift voor Europees recht*, nr. 5/6: 132-140.

Kulk, S. en S. van Deursen (2020) *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, Den Haag: Wetenschappelijk Onderzoek- en Documentatiecentrum.

Kurzweil, R. (2005) *The Singularity is Near: When Humans Transcend Biology*, Londen: Penguin.

Lancieri, F. en P. Sakowski (2020) 'Competition in Digital Markets: A Review of Expert Reports', *Stanford Journal of Law, Business & Finance*, 26: 65.

Lawton, G. (2019) 'AI can predict your future behaviour with powerful new simulations; *New Scientist*, 2 oktober 2019. Beschikbaar op: https://www.newscientist.com/article/mg24332500-800-ai-can-predict-your-future-behaviour-with-powerful-new-simulations/

Le Maire, B., P. Altmaier en M. Keijzer (2021) *Strengthening the Digital Markets Act and Its Enforcement*, non-paper DMA. Beschikbaar op: https://www.rijksoverheid.nl/documenten/publicaties/2021/05/26/non-paper-dma

Lee, K. F. (2018) *AI Superpowers: China, Silicon Valley, and the New World Order*, Boston: Houghton Mifflin Harcourt.

Lee, R. en S. Vaughan (2010) 'REACHING down: Nanomaterials and chemical safety in the European Union', *Law, Innovation and Technology,* 2(2): 193-217.

Leeuw, F. (2020) *Van Legal Realism naar Legal Big Data: Ontwikkelingen in empirisch-juridisch onderzoek toen, nu en straks*, Den Haag: Boom Juridisch.

Leonard, M. (red.) (2016) *Connectivity Wars: Why migration, finance and trade are the geo-economic battlegrounds of the future,* Londen: European Council on Foreign Relations.

Lessig, L. (2006) *Code and other laws of cyberspace, version 2.0*, New York: Basic Books.

Leung, J. (2019) *Who will govern artificial intelligence? Learning from the history of strategic politics in emerging technologies*, dissertatie, Oxford: Oxford University. Beschikbaar op: https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665

Lewis, P. en P. Hilder (2018) 'Leaked: Cambridge Analytica's blueprint for Trump victory', *The Guardian*, 23 maart 2018. Beschikbaar op: https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory

Lewis-Kraus, G. (2016) 'The great A.I. awakening', *The New York Times Magazine*. Beschikbaar op: https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html

Libicki, M. (2019) 'A Hacker Way of Warfare', blz. 137-142 in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, Montgomery: Air University Press.

Lin, H. (2019) 'Escalation Risk in an Artificial Intelligence-Infused World', blz. 143-152 in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, Montgomery: Air University Press.

Liu, X., L. Faes, A. Kale, S. Wagner, D.J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern, J. Ledsam, M. Schmid, K. Balaskas, E. Topol, L. Bachmann, P. Keane en A. Denniston (2019) 'A comparison of deep

learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis', *The Lancet Digital Health*, 1: e271-297.

Loucks, J., S. Hupfer, D. Jarvis en T. Murphy (2019) *Future in the balance? How countries are pursuing an AI advantage*, Deloitte Center for Technology, Media & Telecommunications. Beschikbaar op: https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/public-sector/lu-global-ai-survey.pdf

Luttwak, E. (1990) 'From geopolitics to geo-economics: Logic of conflict, grammar of commerce', *The National Interest*, 20: 17-23.

Lynch, S. (2021) *Andrew Ng: Why AI Is the New Electricity*, Stanford Graduate School of Business, 4 mei 2021. Beschikbaar op: https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity

Macaulay, T. (2020) 'The Guardian's GPT-3-generated article is everything wrong with AI media hype', *The Next Web*, 8 september 2020. Beschikbaar op: https://thenextweb.com/news/the-guardians-gpt-3-generated-article-is-everything-wrong-with-ai-media-hype

Marchant, G. (2011) 'The Growing Gap Between Emerging Technologies and the Law', blz. 19-33 in G.E. Marchant, B. Allenby en J. Herkert (reds.) *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight The Pacing Problem*, Leiden: Springer.

Marcus, G. (2018) *Deep learning: A critical appraisal*. Beschikbaar op: https://arxiv.org/pdf/1801.00631.pdf?ut

Marcus, G. en E. Davi, (2019) *Rebooting AI: Building Artificial Intelligence We Can Trust*, New York: Vintage.

Marsh, H. (2019) 'Can man ever build a mind?', *Financial Times*, 10 januari 2019. Beschikbaar op: https://www.ft.com/content/2e75c04a-0f43-11e9-acdc-4d9976f1533b

Marx, K. en F. Engels (2010 [1932]) *Die deutsche ideologie (Vol. 17),* Berlin: Akademie Verlag.

Mateescu, A. en A. Ngyun (2019) *Explainer: Workplace Monitoring en Surveillance,* New York: Data en Society Research Institute. Beschikbaar op: https://datasociety.net/wp-content/uploads/2019/02/DS_Workplace_Monitoring_Surveillance_Explainer.pdf

May, T. (1994) *The Cyphernomicon: Cypherpunks FAQ and More, Version 0.666.* Beschikbaar op: https://hackmd.io/@jmsjsph/TheCyphernomicon

Mayor, A. (2018) *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology*, Princeton, VS: Princeton University Press.

Mazzucato, M. (2014) *The Entrepreneurial State: Debunking Public vs. Private Myths*, Londen: Anthem Press.

McCulloch, W. en W. Pitts (1943) 'A logical calculus of the ideas immanent in nervous activity', *The Bulletin of Mathematical Biophysics,* 5(4): 115-133.

McKinsey & Company (2020) *How nine digital front-runners can lead on AI in Europe*, McKinsey & Company. Beschikbaar op: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/how%20nine%20digital%20frontrunners%20can%20lead%20on%20ai%20in%20europe/how-nine-digital-frontrunners-can-lead-on-ai-in-europe.pdf

McLuhan, M. (1994 [1964]) *Understanding Media. The Extensions of Man,* Cambridge: MIT Press.

Meijer, A., S. Grimmelikhuijsen en M. Bovens (2021) 'De legitimiteit van het algoritmisch bestuur', *Nederlands Juristenblad*, 18: 1470-1478.

Menting, M.C. (2016) *Industry Codes of Conduct in a Multi-layered Dutch Private Law*, dissertatie, Tilburg: Tilburg University.

Ministerie van Defensie (2020) *Defensievisie 2035*, Den Haag: Ministerie van Defensie.

Ministerie van Economische Zaken en Klimaat (2018) *Nederlandse Digitaliseringsstrategie*, Den Haag: Ministerie van Economische Zaken en Klimaat. Beschikbaar op: https://www.rijksoverheid.nl/documenten/kamerstukken/2021/04/26/nederlandse-digitaliseringsstrategie-2021

Ministerie van Economische Zaken en Klimaat, Ministerie van Justitie en Veiligheid, Ministerie van Sociale Zaken en Werkgelegenheid, Ministerie van Onderwijs, Cultuur en Wetenschap, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2019) *Strategisch Actieplan voor AI*, bijlage bij Kamerstukken 26 643 en 32 761, nr. 640, Den Haag: Ministerie van Economische Zaken. Beschikbaar op: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/beleidsnotas/2019/10/08/strategisch-actieplan-voor-artificiele-intelligentie/Rapport+SAPAI.pdf

Misuraca, G. en C. van Noordt (2020) *AI Watch - Artificial Intelligence in public services: Overview of the use and impact of* AI *in public services in the* EU, Luxemburg: Publications Office of the European Union.

Mitchell, M. (2021) *Why AI is harder than we think*. Beschikbaar op: https://arxiv.org/pdf/2104.12871.pdf

Mittelstadt, B. (2019) 'Principles alone cannot guarantee ethical AI', *Nature Machine Intelligence*, 1: 501-507.

Moerel, L. en C. Prins (2016) 'Privacy voor de homo digitalis: Proeve van een nieuw toetsingskader voor gegevensbescherming in het licht van Big Data en Internet of Things', blz. 9-124 in *Homo Digitalis, Preadviezen 2016 Nederlandse Juristen-Vereniging 2016*, Alphen aan den Rijn: Kluwer.

Moerel, L. en C. Prins (2016) *Privacy for the Homo Digitalis: Proposal for a New Regulatory Framework for Data Protection in the Light of Big Data and the Internet of Things*, Wolters Kluwer. Beschikbaar op: https://dx.doi.org/10.2139/ssrn.2784123

Moore, M. en D. Tambini (reds.) (2018) *Digital dominance: power of Google, Amazon, Facebook, and Apple*, Oxford: Oxford University Press.

Moravec, H. (1988) *Mind Children: The Future of Robot and Human Intelligence*, Cambridge: Harvard University Press.
Morgus, R. (2019) 'The Spread of Russia's Digital Authoritarianism', blz. 89-97 in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, Montgomery: Air University Press.
Morozov, E. (2011) *The Net Delusion: How to Not Liberate the World*, Londen: Penguin.
Morozov, E. (2013) *To Save Everything, Click Here: The Folly of Technological Solutionism*, Londen: Penguin.
Mozur, P. (2019) 'One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority', *The New York Times*, 14 april 2019. Beschikbaar op: https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html
National Security Commission on Artificial Intelligence (2021) *Final Report*, Arlington: NSCAI.
Nationale Proeftuin Precisie Landbouw (z.d.) *Precisielandbouw voor alle telers*, Nationale Proeftuin Precisie Landbouw. Beschikbaar op: https://www.proeftuinprecisielandbouw.nl/
NAVO (2019) *Performance Audit Reports International Board of Auditors for nato (iban)*. Beschikbaar op: www.nato.int/cps/en/natolive/topics_111783.htm
Nemitz, P. (2018) 'Constitutional Democracy and Technology in the age of Artificial Intelligence', *Philosophical Transactions of the Royal Society A*:, 376(2133), 20180089. Beschikbaar op: https://doi.org/10.1098/rsta.2018.0089
Nemitz, P. en M. Pfeffer (2020) *Prinzip mensch. Macht, Freiheit und Democratie im Zeitalter der Künstlichen Intelligenz*, Bonn: Dietz.
Neumann, J. von (2012 [1958]) *The Computer and the Brain*, New Haven: Yale University Press.
Nilsson, N. (2009) *The quest for artificial intelligence*, Cambridge: Cambridge University Press.
Noble, S. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York: New York University Press.
Noort, W. van (2018) 'Nederland kampt met braindrain in artificiële intelligentie', *NRC*, 27 augustus 2018. Beschikbaar op: https://www.nrc.nl/nieuws/2018/08/27/nederland-kampt-met-ai-braindrain-a1614393
NOS (2018) *AI-special: het spanningsvlak tussen mens en machine*, NOS, 1 oktober 2018. Beschikbaar op: https://nos.nl/nieuwsuur/artikel/2252834-ai-special-het-spanningsvlak-tussen-mens-en-machine
NOS (2016) *Stiekem experiment op Facebook*, NOS, 29 juni 2016. Beschikbaar op: https://nos.nl/artikel/668173-stiekem-experiment-op-facebook

Nuvolari, A. (2019) 'Understanding successive industrial revolutions: A "development block" approach', *Environmental Innovation and Societal Transitions,* 32: 33-44.
OESO (2018) *Private Equity Investment in Artificial Intelligence,* OECD Going Digital Policy Note, Parijs: OECD Publishing. Beschikbaar op: www.oecd.org/going-digital/ai/private-equity-investment-in-artificial-intelligence.pdf
OESO (2019) *Artificial Intelligence in Society*, Parijs: OECD Publishing. Beschikbaar op: https://doi.org/10.1787/eedfee77-en
OESO (z.d.) *What are the oecd Principles on AI?* Beschikbaar op: https://www.oecd.org/going-digital/ai/principles/
Olsthoorn, P. (2015) *25 jaar internet in Nederland*, Amsterdam: Fast Moving Targets.
Onderwijsraad (2017) *Doordacht digitaal: Onderwijs in het digitale tijdperk,* Den Haag: Onderwijsraad. Beschikbaar op: https://www.onderwijsraad.nl/upload/documents/publicaties/volledig/Doordacht digitaal-a.pdf
O'Neil, C. (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*, Londen: Penguin.
OneThird (z.d.) *Our Food Loss and Waste Solutions*. Beschikbaar op: https://onethird.io/food-waste-solutions/
Overheidsbreed Beleidsoverleg Digitale Overheid (2018) *NL DIGIbeter: Agenda Digitale Overheid*, Digitale Overheid. Beschikbaar op: https://www.digitaleoverheid.nl/wp-content/uploads/sites/8/2018/07/nl-digibeter-agenda-digitale-overheid.pdf
OvV (2019) *Wie stuurt? Verkeersveiligheid en automatisering in het wegverkeer*, Den Haag: Onderzoeksraad voor Veiligheid. Beschikbaar op: https://www.onderzoeksraad.nl/nl/media/attachment/2019/11/28/wie_stuurt_verkeersveiligheid_en_automatisering_in_het_wegverkeer.pdf
Pall, M. (2019) *How the Telecommunications Industry 5G Strategy Will Use Artificial Intelligence to Replace Human Intelligence: The End of Mankind as We Know It*, 8 juni 2019. Beschikbaar op: www.salzburg.gv.at/gesundheit_/Documents/5G-AI%20(002).pdf
Pasquale, F. (2015) *Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge: Harvard University Press.
Pasquale, F. (2020) *New Laws of Robotics: Defending Human Expertise in the Age of AI,* Cambridge: Harvard University Press.
Passchier, R. (2020) *Artificiële intelligentie en de rechtsstaat. Over verschuivende overheidsmacht, Big Tech en de noodzaak van constitutioneel onderhoud*, Amsterdam: Boom.
Pearl, J. (2019) 'The limitations of opaque learning machines', blz. 13-19 in J. Brockman (red.) *Possible Minds: Twenty-Five Ways of Looking at AI*, Londen: Penguin.

Perez, C. (2003) *Technological revolutions and financial capital: The Dynamics of Bubbles and Golden Ages*, Cheltenham: Edward Elgar Publishing.
Perez, C. (2016) 'Capitalism, technology and a green global golden age: The role of history in helping to shape the future', blz. 191-217 in M. Jacobs en M. Mazzucato (reds.) *Rethinking Capitalism: Economics and policy for sustainable and inclusive growth*, Chichester: John Wiley en Sons.
Perez, C. (2017-2020) *Second Machine Age or Fifth Technological Revolution?*, blog. Beschikbaar op: http://beyondthetechrevolution.com/blog/
Perez, C.C. (2019) *Invisible Women: Exposing Data Bias in a World Designed for Men*, New York: Vintage.
Perrault, R., Y. Shoham, E. Brynjolfsson, J. Clark, J. Etchemendy, B. Grosz, T. Lyons, J. Manyika, S. Mishra en J. Niebles (2019) *The AI Index 2019 Annual Report*, Stanford: Stanford University, Human-Centered AI Institute. Beschikbaar op: https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf
Pethokoukis, J. (2019) 'How AI is like that other general purpose technology, electricity', blog, *AEI*, 25 november 2019. Beschikbaar op: https://www.aei.org/economics/how-ai-is-like-that-other-general-purpose-technology-electricity/
PBL (2017) *Mobiliteit en elektriciteit in het digitale tijdperk. Publieke waarden onder spanning*, Den Haag: Planbureau voor de Leefomgeving. Beschikbaar op: https://www.pbl.nl/sites/default/files/downloads/pbl-2017-mobiliteit-en-elektriciteit-in-het-digitale-tijdperk-1874.pdf
Politie (2018) *Nieuwe technologie in oude politiezaken*, 23 mei 2018. Beschikbaar op: https://www.politie.nl/nieuws/2018/mei/23/00-nieuwe-technologie-in-oude-politiezaken.html
Polyakova, A. en C. Meserole (2019) *Exporting digital authoritarianism: The Russian and Chinese models*, Policy Brief, Democracy and Disorder Series, Washington: Brookings.
Poon, M. (2016) 'Corporate Capitalism and the Growing Power of Big Data: Review Essay', *Science, Technology, & Human Values*, 41:1088-1108.
Prins, C. (2017) 'Politiek profileren', *Nederlands Juristenblad*, 92(38): 2799.
Prins, C. (2018) 'Urgenda en Digitalisering', *Nederlands Juristenblad*, 2018/1098, 22.
Prufer, J. en C. Schottmüller (2017) *Competing with Big Data*, TILEC Discussion Paper Nr. 2017-006, CentER Discussion Paper Nr. 2017-007. Beschikbaar op: https://ssrn.com/abstract=2918726 or http://dx.doi.org/10.2139/ssrn.2918726
Pruis, M. (2017) 'Kennen of gekend worden', *De Groene Amsterdammer*, 11 oktober 2017. Beschikbaar op: https://www.groene.nl/artikel/kennen-of-gekend-worden
Purtova, N. (2018) 'The Law of Everything. Broad concept of personal and future of EU Data Protection Law', *Law, Innovation and Technology*, 1: 40-81.

QuTech (2019) *Creating the Quantum Future. QuTech Annual Report 2019*. Beschikbaar op: https://qutech.h5mag.com/annual_report_2019/

Raad van Europa (z.d.) *CAHAI – Ad Hoc Committee on Artificial Intelligence*, Straatsburg: Raad van Europa. Beschikbaar op: https://www.coe.int/en/web/artificial-intelligence/cahai

Raad van State (2018) *Ongevraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen*, Den Haag: Raad van State. W04.18.0230/I. Beschikbaar op: https://www.raadvanstate.nl/@112661/w04-18-0230/

Raad van State (2020) *Jaarverslag 2020*, Den Haag: Raad van State

Raad van State (2020b) *Ongevraagd advies over ministeriële verantwoordelijkheid*, Den Haag: Raad van State. Beschikbaar op: https://www.raadvanstate.nl/@121396/advies-ministeriele-verantwoordelijkheid/

Raad van State (2021) *Digitalisering. Wetgeving en bestuursrechtspraak*, Den Haag: Raad van State. Beschikbaar op: https://www.raadvanstate.nl/publish/library/13/digitalisering_wetgeving_en_bestuursrechtspraak.pdf

Rli (2021) *Digitaal Duurzaam*, Den Haag: Raad voor de leefomgeving en infrastructuur. Beschikbaar op: https://www.rli.nl/sites/default/files/rli_2021-02_digitaal_duurzaam_-_defintiief_advies.pdf

Rechtstreeks (2019) *Algoritmes in de rechtspraak. Wat artificiële intelligentie kan betekenen voor de rechtspraak*, Rechtstreeks 2019, nr. 2, Den Haag: Sdu. Beschikbaar op: https://www.rechtspraak.nl/SiteCollectionDocuments/rechtstreeks-2019-02.pdf

RVS (2019) *Waarde(n)volle zorgtechnologie. Een verkennend advies over de kansen en risico's van kunstmatige intelligentie in de zorg*, Den Haag: Raad voor Volksgezondheid en Samenleving.

ROB (2021) *Sturen of gestuurd worden? Over de legitimiteit van sturen met data*, Den Haag: Raad voor het Openbaar Bestuur. Beschikbaar op: https://www.raadopenbaarbestuur.nl/binaries/raad-openbaar-bestuur/documenten/publicaties/2021/05/25/advies-sturen-of-gestuurd-worden/Sturen_of_gestuurd_worden_Adviesrapport_2021_05.pdf

Ram, A. (2019) 'Tencent trials AI diagnosis program for Parkinson's in London', *Financial Times*, 7 mei 2019. Beschikbaar op: https://www.ft.com/content/183c412a-6766-11e9-9adc-98bf1d35a056

Rao, A. en G. Verweij (2017) *Sizing the Prize: What's the Real Value of AI for Your Business and How Can You Capitalise?*, PricewaterhouseCoopers. Beschikbaar op: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

Rathenau Instituut (2021) *Waardevol gebruik van menselijke DNA-data. Onderzoek naar het borgen van publieke waarden in de waardeketen van DNA-data*, Den Haag: Rathenau Instituut. Beschikbaar op: https://www.rathenau.nl/sites/default/files/2021-05/Waardevol_gebruik_van_menselijke_DNA%20data_Rathenau_Instituut.pdf

Rathenau Instituut (2021) *International mobility of AI scientists*, Factsheet, 7 juni 2021. Beschikbaar op: https://www.rathenau.nl/en/science-figures/international-mobility-ai-scientists

Rechtbank Amsterdam (2020) ECLI:NL:RBAMS:2020:2917 – Rechtbank Amsterdam, 11-06-2020 / C/13/684665 / KG ZA 20-481, uitspraak 11 juni 2020. Beschikbaar op: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBAMS:2020:2917

Rechtbank Amsterdam (2021) ECLI:NL:RBAMS:2021:1018, uitspraak 11 maart 2021. *Rechtspraak.nl*. Beschikbaar op: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBAMS:2021:1018

Rechtbank Amsterdam (2019) ECLI:NL:RBAMS:2019:4799, uitspraak 4 juli 2019. Beschikbaar op: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBAMS:2019:4799

Rechtbank Arnhem (2008) ECLI:NL:RBARN:2008:BD7578, uitspraak 18 juli 2008. Beschikbaar op: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBARN:2008:BD7578

Reclaim Your Face (2021) *61 MEPS urge the EU to ban biometric mass surveillance!*, 16 april 2021. Beschikbaar op: https://reclaimyourface.eu/61-meps-urge-eu-ban-biometric-mass-surveillance/

Reed, C. (2018) 'How should we regulate artificial intelligence?', *Philosophical Transactions of the Royal Society A*, *376*: 20170360. Beschikbaar op: http://dx.doi.org/10.1098/rsta.2017.0360

Reinhold, F. (2021) 'AlgorithmWatch's response to the European Commission's proposed regulation on Artificial Intelligence – A major step with major gaps', *AlgorithmWatch*, 22 april 2021. Beschikbaar op: https://algorithmwatch.org/en/response-to-eu-ai-regulation-proposal-2021/

Reuters (2020) 'False claim: COVID-19 stands for Certification of Vaccination Identification by Artificial Intelligence', *Reuters*, 24 april 2020. Beschikbaar op: https://www.reuters.com/article/uk-factcheck-covid-name-abbreviation/false-claim-covid-19-stands-for-certification-of-vaccination-identification-by-artificial-intelligence-idUSKCN2262AS?edition-redirect=in

Rid, T. (2016) *Rise of the Machines: A Cybernetic History*, New York: WW Norton & Company.

Rijksoverheid (2021) *Minister Dekker maakt kennis met de digitale assistent Julo op de website van het Juridisch Loket*, video, 12 januari 2021. Beschikbaar op: https://www.rijksoverheid.nl/documenten/videos/2021/01/12/minister-dekker-maakt-kennis-met-de-digitale-assistent-julo-op-de-website-van-het-juridisch-loket

Rijksoverheid (2019) *Minister Ollongren komt met maatregelen voor vernieuwing van de democratie*, nieuwsbericht 26 juni 2019. Beschikbaar op: https://www.rijksoverheid.nl/actueel/nieuws/2019/06/26/minister-ollongren-komt-met-maatregelen-voor-vernieuwing-van-de-democratie

Rijksoverheid (2021) *Duitsland, Frankrijk, Nederland: 'alle fusies en overnames digitale poortwachters beoordelen'*, nieuwsbericht 27 mei 2021. Beschikbaar op: https://www.rijksoverheid.nl/actueel/nieuws/2021/05/27/duitsland-frankrijk-nederland-alle-fusies-en-overnames-digitale-poortwachters-beoordelen

Robotique et Mathématiques (2017) 'Kiva Robots: Amazon', *YouTube*, 24 juli 2017. Beschikbaar op: https://www.youtube.com/watch?v=ULswQgd73Tc

Roy, V. van, F. Rossetti, K. Perset en L. Galindo-Romero (2021) *AI Watch - National Strategies on Artificial Intelligence: A European Perspective*, EUR 30745, Luxemburg: Publications Office of the European Union.

Rühlig, T. (2020) *Technical standardisation, China and the future international order: A European perspective*, Brussel: Heinrich Böll Stiftung.

Russell, S. (2019) *Human Compatible: Artificial Intelligence and the Problem of Control*, Londen: Penguin.

Russell, S. en P. Norvig (2020) *Artificial intelligence: a modern approach*, 4de ed., Essex: Pearson.

Rutkin, A. (2014) 'Even online, emotions can be contagious', *New Scientist*, 25 juni. Beschikbaar op: https://www.newscientist.com/article/mg22229754-900-even-online-emotions-can-be-contagious/?ignored=irrelevant#.U7EHbI21au8

Salian, I. (2019) 'AI in the sky aids feet on the ground spotting human rights violations', blog, *NVIDIA*, 4 april 2019. Beschikbaar op: https://blogs.nvidia.com/blog/2019/04/04/human-rights-watch-ai-gtc/#:~:text=AI%20in%20the%20Sky%20Aids%20Feet%20on%20the%20Ground%20Spotting%20Human%20Rights%20Violations&text=In%20a%20traditional%20human%20rights,collect%20hospital%20or%20autopsy%20records.

Sample, I. (2019) 'Human Compatible by Stuart Russell review – AI and our future', *The Guardian*, 24 oktober 2019. Beschikbaar op: https://www.theguardian.com/books/2019/oct/24/human-compatible-ai-problem-control-stuart-russell-review

Scharre, P. (2018) *Army of None: Autonomous Weapons and the Future of War*, New York: WW Norton & Company.

Schick, N. (2020) *Deep Fakes and the Infocalypse: What You Urgently Need to Know*, Londen: Octopus Publishing Group.

Schiphol (2019) 'Schiphol launches pilot for boarding by means of facial recognition', *Schiphol Nieuws*, 18 februari 2019. Beschikbaar op: https://news. schiphol.com/schiphol-launches-pilot-for-boarding-by-means-of-facial-recognition/?

Scholvin, S. en M. Wigell (2018) *Geo-economics as a concept and practice in international relations: surveying the state of the art*, Working Paper nr. 102, Helsinki: Finnish Institute of International Affairs.

Schothorst, Y. en D. Verhue (2018) *Nederlanders over Artificiële Intelligentie. Onderzoek naar de kennis en houding van burgers en ondernemers over Artificiële Intelligentie*, Kantar Public. Beschikbaar op: file://srvr-maz04/Users$/MAZ0846/Downloads/H5851+rapport+AI%20(2).pdf

Schubert, C. (2013) 'How to evaluate creative destruction: reconstructing Schumpeter's approach', *Cambridge Journal of Economics,* 37(2): 227-250.

Schulz, W. en J. van Hoboken (2016) *Human rights and encryption*, Parijs: UNESCO Publishing.

Schuyt, K. (2006) *Steunberen van de samenleving*, Amsterdam: Amsterdam University Press.

Schwab, K. (2016) *The Fourth Industrial Revolution*, New York: Random House.

Scott, C. (2007) 'Rethinking regulatory governance for the age of biotechnology', blz. 19-35 H. Somsen (red.) *The Regulatory Challenge of Biotechnology: Human Genetics, Food and Patents*, Cheltenham: Edward Elgar Publishing.

Select Committee on Artificial Intelligence (2018) *AI in the UK: Ready, Willing and Able?*, Report of Session 2017-19, HL Paper 100, Londen: Authority of the House of Lords. Beschikbaar op: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

Semaan, N. (2020, 16 maart) Die Demokratisierung von Deepfakes: Wie technologische Entwicklung unseren gesellschaftlichen Konsens beeinflussen kann, *International Reports*, 1: 60-68.

Seo, S. (2019) *Policing the Open Road: How cars transformed American freedom*, Cambridge: Harvard University Press.

SER (2016) *Mens en technologie, samen aan het werk*, Den Haag: Sociaal-Economische Raad. Beschikbaar op: https://www.ser.nl/-/media/ser/downloads/adviezen/2016/mens-technologie-publieksversie.pdf

SER (2018) *Technologische ontwikkelingen en rol ondernemingsraad. Handreiking voor ondernemingsraden*, Den Haag: Sociaal-Economische Raad.

Sheikh, H. (2021) Aanbevelingen. ESB, 106(4801), 407-410. Beschikbaar op: https://esb.nu/esb/20066595/aanbevelingen-voor-een-geo-economische-wereld

Sheikh, H. en P. Timmers (2020) 'Na Trump is het tijd voor 'Make Europe Great Again'', *NRC*, 3 december. Beschikbaar op: https://www.nrc.nl/nieuws/2020/12/03/na-trump-is-het-tijd-voor-make-europe-great-again-a4022477

Simonite, T. (2019) 'Behind the Rise of China's Facial-Recognition Giants', *Wired*, 3 september 2019. Beschikbaar op: https://www.wired.com/story/behind-rise-chinas-facial-recognition-giants/

Singer, P. en E. Brooking (2018) *LikeWar: The Weaponization of Social Media*, Boston: Mariner Books.

SkinVision (z.d.) *Ontdek de slimme huidcheck*. Beschikbaar op: https://www.skinvision.com/nl/

Sloot, B. van der en van S. Schendel (2020) 'Tien voorstellen voor aanpassingen aan het Nederlands procesrecht in het licht van Big Data', *Computerrecht*, 1: 4-13.

Sloot, B. van der, E. Keymolen, M. Noorman, M. Pechenizkiy, H. Weerts, Y. Wagensveld, B. Visser, i.s.m. het College voor de Rechten van de Mens (2021) *Non-discriminatie by design*, Tilburg: Tilburg University. Beschikbaar op: www.tilburguniversity.edu/sites/default/files/download/01%20handreiking%20non-discriminatie%20by%20design%28NL%29.pdf

Smith, C. (2013) 'Facebook Users Are Uploading 350 Million New Photos Each Day', *Business Insider*, 18 september 2013. Beschikbaar op: https://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9?IR=T

Smith, C. (1999) 'International collaboration in science and technology: lessons from CERN', *European Review*, 7(1): 77-92.

Smits, J. (2015) Wetgeving en andere normenstelsels: zes aanwijzingen aan de Nederlandse wetgever. *RegelMaat 30*(5), 357-359.

Smuha, N. (2019) 'From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence', *Law, Innovation and Technology*, 13(1): 57-84.

Smuha, N., E. Ahmed-Rengers, A. Harkens, W. Li, J. MacLaren, R. Piselli en K. Yeung (2021) *How the EU can achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act*. Beschikbaar op: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3899991_code3594902.pdf?abstractid=3899991&mirid=1

Soldatov, A. en I. Borogan (2015) *The Red Web: The Struggle Between Russia's Digital Dictators and the New Online Revolutionaries*, New York: Public Affairs.

Solove, D. (2011) *Nothing to hide. The False Tradeoff Between Privacy and Security,* New Haven: Yale University Press.

Šopova, J. (2018) 'Audrey Azoulay: Making the most of artificial intelligence', *The UNESCO Courier*, 3: 36-41. Beschikbaar op: https://unesdoc.unesco.org/ark:/48223/pf0000265211

Staatscourant (2009) 'Convenant tussen de Sociale Inlichtingen- en Opsporingsdienst en de Stichting Inlichtingenbureau', *Staatscourant van het Koninkrijk der Nederlanden 2009-11*, 19 januari 2009. Beschikbaar op: https://zoek.officielebekendmakingen.nl/stcrt-2009-791.html

Staatscourant-2017-69426 (2017) *Besluit van de Minister-President, Minister van Algemene Zaken, van 22 december 2017, nr. 3215945, houdende vaststelling van de tiende wijziging van de Aanwijzingen voor de regelgeving*, Den Haag: Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

Staatscourant-2017-69426 (2017) *Besluit van de Minister-President, Minister van Algemene Zaken, van 22 december 2017, nr. 3215945, houdende vaststelling*

van de tiende wijziging van de Aanwijzingen voor de regelgeving, Den Haag: Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.
Steijns, M. (2021) "Van repliek gediend?" Een verkenning van tegenmacht vanuit maatschappelijke organisaties. WRR Working Paper nr. 50, Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid.
Stikker, M. (2019) *Het internet is stuk*, Amsterdam: De Geus.
SURF (2021) 'SURF start met bouw nieuwe nationale supercomputer', *SURF*, 8 februari. Beschikbaar op: https://www.surf.nl/nieuws/surf-start-met-bouw-nieuwe-nationale-supercomputer
Svantesson, D. (2020) 'Is International Law Ready for the (Already Ongoing) Digital Age? Perspectives from Private and Public International Law', blz. 113-155 in M. Busstra, W. Theeuwen, Y. Buruma & D. Svantesson (reds.) *International Law for a Digitalised World,* Royal Netherlands Society of International Law, Collected Papers nr. 147, Den Haag: T.M.C. Asser Press.
Sykes, K. en P. Macnaghtan (2013) 'Responsible innovation: opening up dialog and debate', blz. 85-107 in R. Owen, J. Bessant en M. Heintz (reds.) *Responsible innovation: Managing the responsible emergence of science and innovation in society*, Chichester: John Wiley & Sons.
Taplin, J. (2017) *Move Fast and Break Things: How Facebook, Google, and Amazon have cornered culture and what it means for all of us*, New York: Pan Macmillan.
Taylor, L., L. Floridi en B. van der Sloot (reds.) (2016) *Group privacy: New challenges of data technologies*, Dordrecht: Springer.
Tegmark, M. (2017) *Life 3.0: Being Human in the Age of Artificial Intelligence*, Londen: Penguin.
Tenner, E. (1997) *Why Things Bite Back: Technology and the revenge of unintended consequences*, New York: Vintage.
Thiel, P. (2009) 'The Education of a Libertarian', *Cato Unbound*, 13 april 2009. Beschikbaar op: https://www.cato-unbound.org/2009/04/13/peter-thiel/education-libertarian
Thomas, D. (2021)' Is This Beverly Hills Cop Playing Sublime's 'Santeria' to Avoid Being Live-Streamed?', *VICE*, 9 februari 2021. Beschikbaar op: https://www.vice.com/en/article/bvxb94/is-this-beverly-hills-cop-playing-sublimes-santeria-to-avoid-being-livestreamed
Tielbeke, J. (2018) 'Lessen van de Luddieten', *De Groene Amsterdammer*, 16 mei 2018. Beschikbaar op: https://www.groene.nl/artikel/lessen-van-de-luddieten
Tijdelijke Commissie Digitale Toekomst (2020) *Update vereist: naar meer parlementaire grip op digitalisering*, eindrapport, Den Haag: Tweede Kamer der Staten-Generaal. Beschikbaar op: https://www.tweedekamer.nl/sites/default/files/atoms/files/eindrapport_tijdelijke_commissie_digitale_toekomst_tweede_kamer_der_staten-generaal.pdf

Tijdschrift voor Toezicht (2020) Aflevering 1, Boom Juridisch Tijdschriften. Beschikbaar op: https://www.bjutijdschriften.nl/tijdschrift/tijdschrifttoezicht/2020/1

Tilburg University (z.d.) *Zero Hunger Lab: Met wiskunde minder honger in de winter*. Beschikbaar op: https://www.tilburguniversity.edu/nl/onderzoek/impact/creating-value-data/zero-hunger-lab

Tilley, A. (2016) 'Alphabet's 'Moonshots' Head Astro Teller: Fear Of AI And Robots Is Wildly Overblown', *Forbes*, 24 maart 2016. Beschikbaar op: www.forbes.com/sites/aarontilley/2016/03/24/alphabets-moonshots-head-astro-teller-fear-of-ai-and-robots-is-wildly-overblown/?sh=7246137973bb

Timmers, P. (2019) 'Challenged by "Digital Soevereignty"', *Journal of Internet Law*, 23(6): 12-21.

Timmers, P. en F. Dezeure (2021) *Nederlandse strategische autonomie en cybersecurity*, onderzoek in opdracht van de Cyber Security Raad, Den Haag: Cyber Security Raad. Beschikbaar op: https://www.cybersecurityraad.nl/binaries/cybersecurityraad/documenten/rapporten/2021/02/18/onderzoeksrapport-digitale-autonomie/Onderzoeksrapport+%27Nederlandse+strategische+autonomie+en+cybersecurity%27.pdf

TNO (2021) *Het technologische ecosysteem van AI in Nederland*, WRR Working Paper nr. 47, Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid.

Tonin, M. (2019) 'Artificial Intelligence: Implications for NATO's Armed Forces', *149 STCTTS 19 E rev. 1 fin.*

Topol, E. (2019) 'High-performance medicine: the convergence of human and artificial intelligence', *Nature medicine*, 25(1): 44-56.

Trajtenberg, M. (2018) *AI as the next GPT: a Political-Economy Perspective*, NBER Working Paper Series, nr. 24245, Cambridge, MA: National Bureau of Economic Research. Beschikbaar op: https://www.nber.org/system/files/working_papers/w24245/w24245.pdf

Trommel, S. (2021) 'Europese datastrategie vereist nationale strategie', *iBestuur online*, 20 april 2021. Beschikbaar op: https://ibestuur.nl/magazine/europese-datastrategie-vereist-nationale-regie

Trouw (2016) 'RDW vindt 'Autopilot' van Tesla misleidende term', *Trouw*, 17 oktober 2016. Beschikbaar op: www.trouw.nl/nieuws/rdw-vindt-autopilot-van-tesla-misleidende-term~b191a2e3/

TU Delft (z.d.) *De computer als herdershond*. Beschikbaar op: https://www.tudelft.nl/stories/articles/de-computer-als-herdershond

Turing, A. (2009 [1950]) 'Computing Machinery and Intelligence', In: R. Epstein, Roberts, G., Beber, G. (reds.) *Parsing the Turing Test*, Dordrecht: Springer.

Turner, F. (2006) *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*, Chicago: University of Chicago Press.

Tweede Kamer (z.d.) *Vaste commissie digitale zaken*, Den Haag: Tweede Kamer der Staten-Generaal. Beschikbaar op: https://www.tweedekamer.nl/kamerleden_en_commissies/commissies/diza

UNESCO (z.d.) *Elaboration of a Recommendation on the ethics of artificial intelligence*. Beschikbaar op: https://en.unesco.org/artificial-intelligence/ethics

Veenstra, A.F. van, S. Djafari, F. Grommé, B. Kotterink en R. Baartmans (2019) *Quick Scan AI in de Publieke Dienstverlening*, Den Haag: TNO. Beschikbaar op: www.rijksoverheid.nl/documenten/rapporten/2019/04/08/quick-scan-in-de-publiekedienstverlening.

Veale, M. (2020) 'A critical take on the policy recommendations of the EU high-level expert group on artificial intelligence', *European Journal of Risk Regulation*, 1-10.

Veale, M. en F. Zuiderveen Borgesius (2021) 'Demystifying the draft EU artificial intelligence act', *Computer Law Review International*, 22(4) (te verschijnen).

Verbeek, P.P. (2014) *Op de vleugels van Icarus: Hoe techniek en moraal met elkaar meebewegen*, Rotterdam: Lemniscaat.

Verhagen, L. (2021) 'Europa komt als eerste ter wereld met regels voor kunstmatige intelligentie. 'Dit wordt een feest voor juristen'', *de Volkskrant*, 21 april 2021. Beschikbaar op: https://www.volkskrant.nl/nieuws-achtergrond/europa-komt-als-eerste-ter-wereld-met-regels-voor-kunstmatige-intelligentie-dit-wordt-een-feest-voor-juristen~b2509f56/

Verhey, L. en N. Verheij (2005) 'De macht van de marktmeesters: Markttoezicht in constitutioneel perspectief', blz. 135-332 in A.A. Rossum, L.F.M. Verhey en N. Verheij (red.) *Toezicht*, Vol. 135, Handelingen der Nederlandse Juristen-Vereniging, Deventer: Kluwer.

Vermaas, P., D. Nas, L. Vandersypen en D. Elkouss Coronas (2019) *Quantum internet: The internet's next big step*, Delft: TU Delft.

Vetzo, M., J. Gerards en R. Nehmelman (2018) *Algoritmes en grondrechten*, Den Haag: Boom Juridisch.

Villani, C. (2018) *For a Meaningful Artificial Intelligence: Towards a French and European Strategy*. Beschikbaar op: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

Vinge, V. (1993) 'The coming technological singularity: How to survive in the post-human era', blz. 11-22 in *Vision-21. Interdisciplinary Science and Engineering in the Era of Cyberspace*, NASA Conference Publication 10129, Cleveland, OH: NASA Lewis Research Center. Beschikbaar op: https://www.researchgate.net/profile/Carol-Stoker-2/publication/234229828_Telepresence_in_the_human_exploration_of_Mars_Field_studies_in_analog_environments/links/554bb5600cf29752ee7e78f8/Telepresence-in-the-human-exploration-of-Mars-Field-studies-in-analog-environments.pdf#page=23

Vleuten, E. van der, R. Oldenziel en M. Davids (2017) *Engineering the future, understanding the past: A social history of technology*, Amsterdam: Amsterdam University Press.

Voorzieningenrechter Arnhem (2008) *Radboud Universiteit mag artikel Mifare Classic Chip publiceren*, persbericht, 18 juli 2008, Rechtbank Arnhem. Beschikbaar op: https://www.sos.cs.ru.nl/applications/rfid/pressrelease-courtdecision.nl.html

Vorst, T. van der, N. Jelicic, M. de Vries en J. Albers (2019) *De (on)mogelijkheden van kunstmatige intelligentie in het onderwijs*, Nr. 2018.068.1828, Utrecht: Dialogic. Beschikbaar op: https://www.dialogic.nl/wp-content/uploads/2019/04/Dialogic-De-onmogelijkheden-van-kunstmatige-intelligentie-in-het-onderwijs-v1.0.116.pdf

VSNU (2021) *Advies publieke waarden voor het onderwijs*, Den Haag: VSNU. Beschikbaar op: https://vsnu.nl/files/documenten/Domeinen/Onderwijs/Advies_werkgroep_publieke_waarden_onderwijs.pdf

Waag (2020) *Algoritme: de mens in de machine - Casusonderzoek toepasbaarheid van conceptrichtlijnen voor algoritmen*, Amsterdam: Waag.

Waardenburg, L., A. Sergeeva en M. Huysman (2020) 'Predictive policing ontcijferd: Een etnografie van het 'Criminaliteits Anticipatie Systeem' in de praktijk', blz. 69-88 in J. Janssens, W. Broer, M. Crispel en R. Salet (reds.) *Informatiegestuurde politie*, Cahiers Politiestudies 54, 's-Hertogenbosch: Gompel & Svacina.

Waarlo, N. en L. Verhagen (2020) 'De stand van gezichtsherkenning in Nederland', *de Volkskrant*, 27 maart 2020. Beschikbaar op : https://www.volkskrant.nl/kijkverder/v/2020/de-stand-van-gezichtsherkenning-in-nederland%7Ev91028/?referrer=https%3A%2F%2Fwww.google.com%2F

Wade, R. (2018) *Governing the market*, Princeton: Princeton University Press.

Walch, K. (2020) 'Why The Race For AI Dominance Is More Global Than You Think', *Forbes*, 9 februari 2020. Beschikbaar op: https://www.forbes.com/sites/cognitiveworld/2020/02/09/why-the-race-for-ai-dominance-is-more-global-than-you-think/?sh=3a34ad2b121f

Waldrop, M. (2019) 'News Feature: What are the limits of deep learning?', *Proceedings of the National Academy of Sciences*, 116(4): 1074-1077.

Wallace, R. (2021) 'The names have changed, but the game's the same': artificial intelligence and racial policy in the USA', *AI and Ethics*: 1-6.

Wallach, W. (2015) *A dangerous master. How to keep technology from slipping beyond our control*, New York: Basic Books.

Warzel, C. (2018) 'Believable: The Terrifying Future of Fake News', *Buzzfeed News*, 11 februari 2018. Beschikbaar op: buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news

Weber, V. (2019) 'Understanding the Global Ramifications of China's Information-Control Model', blz. 76-80 in N. Wright (red.) *Artificial*

Intelligence, China, Russia and the Global Order, Montgomery: Air University Press.
Weinberger, S. (2019) *The imagineers of war: the untold history of DARPA, the Pentagon agency that changed the world*, New York: Vintage.
Weiser, M. (1991) 'The computer for the 21st century', *IEEE Pervasive Computer,* 1(1): 19-25.
Went, R., M. Kremer en A. Knottnerus (red.) (2015) *De robot de baas. De toekomst van werk in het tweede machinetijdperk*, WRR-Verkenning nr. 31. Amsterdam: Amsterdam University Press.
Wiener, N. (1964) *God and Golem, Inc.: A Comment on Certain Points where Cybernetics Impinges on Religion,* Cambridge: MIT Press.
Wiener (2019 [1965]) *Cybernetics: Or Control and Communication in the Animal and the Machine*, Cambridge: MIT Press.
Wilson, H., P. Daugherty en C. Davenport (2019) 'The future of AI will be about less data, not more', *Harvard Business Review*, 14 januari 2019. Beschikbaar op: https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-data-not-more
Winner, I. (1983) 'Technē and Politeia: The Technical Constitution of Society', blz. 97-111 in P. Durbin en F. Rapp (reds.) *Philosophy and Technology, Boston Studies in the Philosophy of Science*, Dordrecht: Springer.
Wittgenstein, L. (1984) *Tractatus logico-philosophicus. Tagebücher 1914-1916. Philosophische Untersuchungen*, Berlijn: Suhrkamp.
Wojciki, S. (2020) *YouTube at 15: My personal journey and the road ahead*, blog, 14 februari 2020. Beschikbaar op: https://blog.youtube/news-and-events/youtube-at-15-my-personal-journey
Wolswinkel, J. (2019) 'Het algoritme van de Afdeling: de realiteit van complex bestuursrecht', *Ars Aequi*, oktober 2019: 776-785. Beschikbaar op: https://pure.uvt.nl/ws/portalfiles/portal/31738610/AA20190776.pdf
Wolswinkel, J. (2020) *Willekeur of algoritme? Laveren tussen analoog en digitaal bestuursrecht*, Tilburg: Tilburg University
WIPO (2019) WIPO *Technology Trends 2019: Artificial Intelligence*, Geneva: World Intellectual Property Organization.
Wouda, F. en H. Hutink (2019) *Artificial Intelligence in de zorg: begrippen, praktijkvoorbeelden en vraagstukken*, Den Haag: Nictiz. Beschikbaar op: https://www.nictiz.nl/wp-content/uploads/Rapport_artificial_intelligence_in_de_zorg.pdf
Wright, G. (2000) 'Review: 'General Purpose Technologies and Economic Growth' (Helpman, 1998)', *Journal of Economic Literature*, 38(1): 161-162.
Wright, N. (2019a) 'Global Competition', blz. 35-41 in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, Montgomery: Air University Press.
Wright, N. (2019b) 'Artificial Intelligence and Domestic Regimes: Digital Authoritarian, Digital Hybrid, and Digital Democracy', blz. 21-34 in N.

Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, Montgomery: Air University Press.
WRR (2008) *Onzekere veiligheid. Verantwoordelijkheid rond fysieke veiligheid,* Amsterdam: Amsterdam University Press.
WRR (2011) *iOverheid*, Amsterdam: Amsterdam University Press.
WRR (2013) *Naar een lerende economie*, Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid.
WRR (2015) *De publieke kern van het internet. Naar een buitenlands internetbeleid*, Amsterdam: Amsterdam University Press.
WRR (2016) *Big Data in een vrije en veilige samenleving*, Amsterdam: Amsterdam University Press.
WRR (2017) *Veiligheid in een wereld van verbindingen. Een strategische visie op het defensiebeleid*, Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid.
WRR (2019) *Voorbereiden op digitale ontwrichting*, Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid.
WRR (2020) *Het betere werk. De nieuwe maatschappelijke opdracht*, Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid.
Wu, T. (2020) *The curse of bigness. How corporate giants came to rule the world,* Londen: Atlantic Books.
Wynsberghe, A. van (2021) 'Sustainable AI: AI for sustainability and the sustainability of AI', *AI and Ethics*, 1: 1-6.
Yeung, K. en M. Lodge (reds.) (2019) *Algorithmic Regulation*, Oxford: Oxford University Press.
Yu, K., A. Beam en I. Kohane (2018) 'Artificial intelligence in healthcare', *Nature biomedical engineering*, 2(10): 719-731.
Zarkadakis, G. (2015) *In our own image: Will artificial intelligence save or destroy us?*, Londen: Ebury Publishing.
Zhang, W.W. (2012) *The China Wave: Rise of a civilizational state*, New Jersey: World Century Publishing Cooperation.
Zhang, B. en A. Dafoe (2019) *Artificial Intelligence: American Attitudes and Trends*, Oxford: University of Oxford, Center for the Governance of AI, Future of Humanity Institute. Beschikbaar op: https://isps.yale.edu/sites/default/files/files/Zhang_us_public_opinion_report_jan_2019.pdf
Zhang, D., S. Mishra, E. Brynjolfsson, J. Etchemendy, D. Ganguli, B. Grosz, T. Lyons, J. Manyika, J. Niebles, M. Sellitto, Y. Shoham, J. Clark en R. Perrault (2021) *The AI Index 2021 Annual Report*, Stanford: Stanford University, Human-Centered AI Institute. Beschikbaar op: https://arxiv.org/ftp/arxiv/papers/2103/2103.06312.pdf
Zielonka, J. (2008) 'Europe as a global actor: Empire by example?', *International Affairs*, 84(3): 471-484.
Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Londen: Profile books.

Ministerie van Economische Zaken en Klimaat

> Retouradres Postbus 20401 2500 EK Den Haag

Wetenschappelijke Raad voor het Regeringsbeleid
T.a.v. mevrouw prof. mr. J.E.J. Prins
Postbus 20004
2500 EA DEN HAAG

Directoraat-generaal Energie, Telecom & Mededinging
Directie Telecommarkt

Bezoekadres
Bezuidenhoutseweg 73
2594 AC Den Haag

Postadres
Postbus 20401
2500 EK Den Haag

Overheidsidentificatienr
00000001003214369000

T 070 379 8911 (algemeen)
F 070 378 6100 (algemeen)
www.rijksoverheid.nl/ezk

Behandeld door

Datum 11 juni 2018
Betreft Adviesaanvraag kunstmatige intelligentie

Ons kenmerk
DGETM-TM / 18101617

Geachte mevrouw Prins,

Kunstmatige intelligentie (artificial intelligence, hierna 'AI'[1]) is aan een opmars bezig. De komst van big data in combinatie met een enorme toename in rekenkracht heeft geleid tot doorbraken in de ontwikkeling van intelligente systemen. Deze systemen kunnen patronen herkennen in grote hoeveelheden data en zijn steeds meer in staat om menselijk redeneren na te bootsen. Ook kunnen ze een steeds bredere reeks aan taken zelfstandig uitvoeren of mensen hierbij assisteren. AI heeft potentieel een grote impact op onze publieke waarden. Om hier meer zicht op te krijgen wil ik de WRR mede namens de Minister van Binnenlandse Zaken en Koninkrijkrelaties, de Staatssecretaris van Binnenlandse Zaken en Koninkrijksrelaties, de Minister van Onderwijs, Cultuur en Wetenschap, de Minister voor Basis- en Voortgezet Onderwijs en Media, de Minister van Volksgezondheid, Welzijn en Sport, de Minister van Justitie en Veiligheid, de Minister voor Rechtsbescherming, de Minister van Landbouw, Natuur en Voedselkwaliteit, de Minister van Defensie, de Minister voor Buitenlandse Handel en Ontwikkelingssamenwerking, de Minister van Buitenlandse Zaken, en de Minister van Infrastructuur en Waterstaat vragen hier nader onderzoek naar te laten doen.

Lopende en eerdere onderzoeken
Er zijn al verschillende onderzoeken uitgevoerd of in gang gezet over de impact van AI.[2] Het ministerie van Binnenlandse Zaken en Koninkrijksrelaties heeft bijvoorbeeld juridisch-wetenschappelijk onderzoek laten verrichten naar de impact van algoritmen op de grondrechten. Ook laat het ministerie van Binnenlandse Zaken en Koninkrijksrelaties onderzoek doen naar (zelflerende) algoritmen die door overheden worden ingezet. Onder begeleiding van het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) wordt ook door het ministerie van

[1] AI heeft als doel om intelligent gedrag na te bootsen en te automatiseren. Voorbeelden van intelligent gedrag zijn kennis vergaren, redeneren, communiceren, problemen oplossen en plannen. Dit gedrag kan in software, machines en apparaten worden geïmplementeerd door middel van technologieën als algoritmes en neurale netwerken. AI is een containerbegrip zonder duidelijke afbakening. Bovendien evolueert het begrip in de tijd.
[2] Het Rathenau Instituut heeft in 2017 twee rapporten gepubliceerd die in dit kader relevant zijn, te weten 'Opwaarderen. Het borgen van publieke waarden in de digitale samenleving' en 'Mensenrechten in het robottijdperk'. Eerder heeft de expertgroep big data en privacy in haar rapport van 2016 aanbevolen om nader onderzoekt te doen naar de impact van AI (Kamerstuk 33009, nr. F). Ook heeft de WRR zelf eerder naar de invloed van big data gekeken in haar rapport 'Big Data in een vrije en veilige samenleving'. Deze rapporten vormden mede aanleiding om meer verdiepend onderzoek te laten doen naar de impact van kunstmatige intelligentie op publieke waarden.

Directoraat-generaal Energie, Telecom & Mededinging
Directie Telecommarkt

Ons kenmerk
DGETM-TM / 18101617

Justitie en Veiligheid onderzoek gedaan naar algoritmen en AI, onder de titel 'Juridische aspecten van algoritmen die zelfstandig besluiten nemen; een verkennend onderzoek'.[3] Het ministerie van Defensie en het ministerie van Buitenlandse Zaken hebben daarnaast eerder advies gevraagd over de toename van autonome functies in wapensystemen.[4] Het ontbreekt echter op dit moment aan een overkoepelend, multidisciplinair onderzoek.

Impact AI

De opmars van AI biedt in de eerste plaats kansen. AI kan bijvoorbeeld worden ingezet om beter te kunnen inschatten wanneer en waar zich bepaalde strafbare feiten kunnen voordoen ("*predictive policing*"), ondersteuning te geven in de rechtspraak, betere doorstroming op de weg mogelijk te maken met zelfrijdende auto's, gezondheidsrisico's te identificeren, ongewenste inhoud op sociale mediaplatforms op te sporen en uitval in het onderwijs te voorkomen.

Ook kan AI bijdragen aan betere besluitvorming binnen zowel de overheid als het bedrijfsleven. Daarnaast kan AI ten goede komen aan het ontwikkelen van nieuwe innovatieve producten en het nemen van weloverwogen aankoopbeslissingen. Bovendien kan de persoonlijke autonomie van burgers worden vergroot doordat zelflerende systemen bijvoorbeeld zorgtaken overnemen en patiënten voor dagelijkse zorg minder afhankelijk worden van anderen. Daarnaast kan AI de menselijke intelligentie aanvullen door de samenwerking tussen mens en machine te faciliteren ('augmented intelligence').

AI brengt echter ook uitdagingen met zich mee voor het borgen van publieke waarden. Zo kan de inzet van AI leiden tot discriminatie. Dit probleem ontstaat bijvoorbeeld als AI-systemen leren met behulp van data die (onbedoeld) bepaalde culturele en sociale ongelijkheden weerspiegelen, welke vervolgens in het systeem worden gereproduceerd. Ook bestaat de vrees dat AI door de analyse van grote hoeveelheden persoonlijke data ongemerkt de keuze(vrijheid) van mensen kan sturen, beïnvloeden en beperken, waardoor de menselijke autonomie vermindert.

Europees en internationaal debat

Op Europees niveau is aandacht voor de impact van AI op publieke waarden en het instrumentarium dat kan worden ingezet om deze waarden te borgen. Zo heeft de Europese Commissie in april 2018 een mededeling over AI gepubliceerd. En, het Parliamentary Assembly van de Raad voor Europa heeft een aanbeveling gepubliceerd ten aanzien van technologische ontwikkelingen, kunstmatige intelligentie en mensenrechten. Daarnaast heeft het Europees Economisch en Sociaal Comité (EESC) zichzelf ten doel gesteld het maatschappelijk debat rondom AI te faciliteren, kennisontwikkeling aan te jagen en een bijdrage te leveren aan het ontwikkelen van Europese beleidskaders voor AI.

Op internationaal niveau staat het onderwerp onder meer binnen de OESO op de agenda. Ook de Verenigde Naties spreken jaarlijks in Geneve over Lethal Autonomous Weapon Systems (LAWS). Bovendien hebben de Verenigde Naties in oktober 2017 een 'Centre for Artificial Intelligence and Robotics' geopend in Den Haag. Dit initiatief brengt de kansen en risico's van AI in kaart, en draagt bij aan een gebalanceerd, internationaal debat. Bovendien hebben een aantal grote technologiebedrijven als Google, Facebook, Amazon, IBM en Microsoft een samenwerkingsverband opgezet waarin de discussie over de maatschappelijke uitdagingen van AI wordt gefaciliteerd.

[3] Deze onderzoeken en onderhavig adviesaanvraag naar de impact van AI geven ook invulling aan het advies van de expertgroep big data en privacy om hier nader onderzoek naar te doen.
[4] Advies van de Adviesraad Internationale Vraagstukken (AIV) en de Commissie van Advies inzake Volkenrechtelijke Vraagstukken (CAVV): 'Autonome wapensystemen - de noodzaak van betekenisvolle menselijke controle'. Dit onderzoek is afgerond.

Directoraat-generaal Energie, Telecom & Mededinging
Directie Telecommarkt

Ons kenmerk
DGETM-TM / 18101617

Verder kan AI een negatieve impact hebben op kwetsbare groepen, bijvoorbeeld lager opgeleiden of burgers met minder geld die zich lastig kunnen wapenen tegen door AI genomen beslissingen. En doordat AI gebruik maakt van veel data, kan dit ook vragen oproepen over de exclusiviteit en integriteit van informatie en de bescherming van persoonsgegevens.

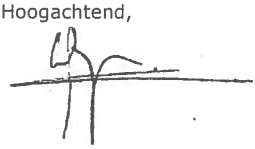
Bovendien zal AI invloed hebben op het terrein van werk. Deze impact kan positief zijn, bijvoorbeeld door het beter bij elkaar kunnen brengen van vraag en aanbod, verbeterde mogelijkheden voor handhaving van wet- en regelgeving en als bron van meer productiviteit. Er kunnen echter ook negatieve gevolgen zijn voor bijvoorbeeld autonomie, werktevredenheid, werkgelegenheid en privacy van werkenden.[5]

Gerelateerd aan bovengenoemde punten speelt de vraag in hoeverre de werking, acties en beslissingen van AI-systemen te begrijpen, controleren en verklaren zijn. En, waar de grenzen liggen aan het gebruik van deze systemen als zij niet transparant en controleerbaar gemaakt kunnen worden. Daarnaast speelt vaak de vraag wie aansprakelijk is voor het (dis)functioneren van (onderdelen van) een AI-systeem. De opkomst van AI versterkt bovendien een aantal bestaande uitdagingen rondom digitalisering, zoals op het gebied privacy, cybersecurity en de macht van een beperkt aantal technologiebedrijven.

Wens voor aanvullend onderzoek
Gelet op het toenemende gebruik van AI in nagenoeg alle sectoren en de hierboven genoemde vragen is er behoefte aan discipline-overstijgend onderzoek naar de impact van AI op publieke waarden. Het kabinet vindt het daarbij van belang te bezien welk instrumentarium reeds voorhanden is om de kansen van AI te faciliteren en de uitdagingen te beantwoorden, en welke instrumenten eventueel aanvullend nodig worden geacht.

Een aantal casestudies kan daarbij nuttig zijn. Een dergelijk onderzoek is niet alleen gewenst en nuttig voor de nationale kennis- en beleidsontwikkeling, maar zal naar verwachting ook richtinggevende inbreng opleveren voor de Europese en internationale discussies over dit onderwerp.

Hoogachtend,



mr. drs. M.C.G. Keijzer
Staatssecretaris van Economische Zaken en Klimaat

[5] Naar de impact van technologie op de arbeidsmarkt en op het karakter van werkzaamheden zijn reeds diverse studies naar verricht, waaronder door de WRR. De vraag is of AI op dit gebied nieuwe vragen oproept, of bestaande vraagstukken extra urgent maakt. Zie bijvoorbeeld de SER verkenning 'Mens en Technologie: samen aan het werk', WRR 'De robot de baas', WRR 'Voor de Zekerheid'.

Rapporten aan de regering

Kiezen voor houdbare zorg. Mensen, middelen en maatschappelijk draagvlak (WRR-rapport nr. 104, 2021)

De Nederlandse zorg presteert over het algemeen goed, maar er zijn in delen van de zorg grote knelpunten. Om kwaliteit en toegankelijkheid te kunnen borgen dient de zorg in financieel, personeel en in maatschappelijk opzicht houdbaar te zijn. Deze drie dimensies van houdbaarheid staan echter steeds meer onder druk door ontwikkelingen als vergrijzing, de opkomst van nieuwe zorgtechnologie en de toename van het aantal chronisch zieken. In dit rapport komt de WRR tot de conclusie dat we – om de groei van de zorg te begrenzen - beter moeten gaan kiezen waar onze prioriteiten in de zorg liggen. Leidend zijn hierbij twee uitgangspunten. Waar kunnen we de meeste gezondheidswinst behalen? En in welke delen van de zorg moeten kwaliteit en toegankelijkheid versterkt worden?

Samenleven in verscheidenheid. Beleid voor de migratiesamenleving (WRR-rapport nr. 103, 2020)

Het is belangrijk dat iedereen – nieuwkomers en gevestigde inwoners – zich thuis kan voelen in Nederland. Dat vraagt om een actiever overheidsbeleid om alle nieuwe migranten wegwijs te maken en op te nemen in onze samenleving. Er dienen ontvangst- en inburgeringsvoorzieningen te komen voor alle migranten: kennismigranten, asielmigranten, gezinsmigranten en migranten uit de Europese Unie. Gemeenten spelen daarin een sleutelrol en hebben daarvoor ondersteuning nodig.

Het betere werk. De nieuwe maatschappelijke opdracht (WRR-rapport nr. 102, 2020)

Nieuwe technologie, de toename van flexibel werk en de intensivering van werk kunnen grote gevolgen hebben voor de kwaliteit van werk. De WRR adviseert bedrijven, instellingen, sociale partners en de overheid in te zetten op goed werk voor iedereen die wil en kan werken. Goed werk is essentieel voor de brede welvaart in ons land: voor de economie en voor de sociale samenhang.

Alle *Rapporten aan de Regering* en publicaties in de reeksen *Policy Briefs*, *Verkenningen* en *Working Papers* zijn beschikbaar via: www.wrr.nl/publicaties.

Opgave AI

De nieuwe systeemtechnologie

Artificiële Intelligentie (AI) is de verbrandingsmotor van de 21e eeuw. De technologie gaat momenteel van het lab de samenleving in en dat roept de vraag op naar de impact ervan op publieke waarden. De WRR biedt in "Opgave AI. De nieuwe systeemtechnologie" hierop een nieuw perspectief. AI valt het beste te vergelijken met de stoommachine, elektriciteit, de verbrandingsmotor en de computer. Dergelijke 'systeemtechnologieën' zijn alomtegenwoordig, kunnen voor allerlei doelen gebruikt worden, en veranderen de economie en samenleving op ingrijpende en onvoorspelbare wijze. We staan momenteel op een keerpunt: AI moet in de samenleving worden ingebed. De overheid in het bijzonder moet daarvoor verschillende opgaven ter hand nemen, zoals het adresseren van onrealistische beelden van AI (demystificatie); het creëren van een goede omgeving om AI te laten werken (contextualisering); het betrekken van maatschappelijke partijen (engagement); het opstellen van een brede wetgevingsagenda voor AI (regulering), en reflectie over de verhouding van Nederland ten opzichte van buitenlandse partijen (positionering).

9 789083 201238 >

WRR

WETENSCHAPPELIJKE RAAD VOOR HET REGERINGSBELEID