# Statement Rondetafelgesprek Drones en Killer Robots

*Koen V. Hindriks, Hoogleraar Kunstmatige Intelligentie, Vrije Universiteit Amsterdam(*), 14-1-2019*

## Kort overzicht huidige stand van zaken in de Kunstmatige Intelligentie

**Kunstmatige Intelligentie (KI) heeft de afgelopen 10-20 jaar enorme vooruitgang geboekt** dankzij de toename van rekenkracht, beschikbare data, en de miniaturisering van kwalitatief hoogwaardige sensoren voor relatief weinig geld. We zijn nu in staat om te redeneren met gigantische hoeveelheden informatie[i], en spraak- en beeldherkenning presteert soms beter dan mensen door nieuwe 'machine learning' technieken[ii]. Met KI zijn we beter dan doctoren in staat kanker te detecteren[iii] en kan Microsoft net zo accuraat als mensen het nieuws automatisch van Chinees naar Engels vertalen[iv].

**Kunstmatige Intelligentie is zeer effectief voor goed omschreven, specifieke taken maar niet voor taken die meer algemeen menselijk inzicht vereisen**. KI presteert goed bij het zoeken in grote datasets, om daar patronen in te herkennen, en is consistent; mensen zijn beter in het maken van common sense afwegingen en het aanvoelen welke factoren relevant zijn voor zo'n afweging. KI presteert vooral beter dan mensen op goed gespecificeerde taken. De generaliseerbaarheid van prestaties in de KI (d.w.z. de prestaties op data die net iets verschillen van de data waarop het systeem getrained is) is nog een van de grootste uitdagingen[v]. Een voorbeeld is een algoritme, dat een object niet meer herkent, als dat wordt geroteerd en teruggeplaatst in dezelfde foto.

**De EU doet nog mee in de 'race om AI' maar investeringen blijven achter bij de VS en China** als we kijken naar de meest recente indicatoren. Een analyse van publicatiecijfers[vi] laat bijvoorbeeld zien dat op dit moment nog de meeste KI papers in de EU worden gepubliceerd (met China sinds een aantal jaren als tweede en de VS als derde) maar dat de impact van de EU papers wel achterblijft op die van de VS[vii]. Om geen grote achterstand op te lopen is het van belang een duidelijke EU visie te ontwikkelen en te investeren in deze sleuteltechnologie[viii]. Ondanks de recente ontwikkelingen zijn er nog grote wetenschappelijke uitdagingen waar Nederlands onderwijs en onderzoek een vooraanstaande rol kan spelen[ix]. In Europa is Nederland helaas een van de landen die nog geen AI strategie heeft geformuleerd, i.t.t. tot Finland, Groot Britannië, Zweden, Frankrijk, Duitsland en België.

## Volledig Geautomatiseerde en Autonome Systemen

**Systemen die kunnen opereren zonder menselijke interventie zijn volledig geautomatiseerd maar daarmee nog niet autonoom**. Drones en robots worden vaak aangeduid als autonome systemen. We kunnen echter beter spreken van (volledig) geautomatiseerde systemen als ze zonder interventie van mensen kunnen opereren[x]. Autonomie in de betekenis van het maken van geïnformeerde beslissingen in een gegeven situatie door alle relevante factoren af te wegen zoals mensen dat kunnen is er nog niet en vereist een veel verder gaande vorm van zelfstandigheid.[xi] Het is bijvoorbeeld niet mogelijk om risicoschattingen van bijkomende schade ('collateral damage') te delegeren aan een machine.

**We zijn nog niet in staat om autonome systemen te maken.** Er kan een analogie van drones met zelfrijdende auto's worden gemaakt. Het is nu mogelijk om zelfrijdende auto's met niveau 2 autonome technologie te ontwikkelen, d.w.z. dat de bestuurder steeds beschikbaar moet zijn om in te grijpen; het ingrijpen van mensen kan echter pas achterwege blijven bij niveau 5. Dat het nog niet mogelijk is om autonome systemen te maken betekent echter niet dat de risico's van het gebruik van moderne technologie zoals drones minder groot zijn. De kans dat er met de huidige (semi-)geautomatiseerde

drones die op grote afstand kunnen worden bestuurd iets misgaat door menselijke inschattingsfouten zal eerder groter zijn, mede omdat drones de operationele beslis-tijd in de praktijk verkorten.

**Kunstmatige Intelligentie is niet in staat de 'rules of engagement' te interpreteren in de context van een operatie.** We zijn niet in staat om autonome systemen te maken die rekening kunnen houden met de verschillende 'rules of engagement' in verschillende contexten (Kroatië vs. Afghanistan bijvoorbeeld). Gegeven de huidige stand van techniek en wetenschap is het ook niet te voorzien wanneer we daartoe wel in staat zullen zijn. Het is daarom noodzakelijk om mensen supervisie te geven en ten allen tijde te blijven betrekken bij het maken van complexe beslissingen (over leven en dood) die we niet kunnen delegeren aan systemen die slechts (volledig) geautomatiseerd zijn. Het is bovendien van belang te zorgen dat mensen voldoende in staat zullen blijven om deze supervisie betekenisvol uit te oefenen, ook met steeds geavanceerdere technologie[xii].

## Drones en Killer Robots

**Drones zijn vliegende camera's die in staat zijn om gedetailleerde informatie te verzamelen**. Drones hebben zeer veel nuttige toepassingen in bijvoorbeeld de landbouw, het reguleren van verkeer, en crisisbeheersing. Drones zijn makkelijk inzetbaar om beelden te verzamelen voor bijvoorbeeld het inschatten van risico's voor mensen. Het is wel van belang ons te blijven realiseren dat automatische herkenning en interpretatie van deze beelden door KI sterk afhankelijk is van hoe deze software is geconfigureerd door mensen (voor welke taak is de KI ontworpen, de mogelijkheid van vooroordelen ('bias') in de interpretatie), evenals de bekende beperkingen zoals gebrek aan generaliseerbaarheid en kwetsbaarheid voor misleiding van deze systemen[xiii].

**Drones zijn toegankelijk geworden en iedereen kan ermee vliegen**. De geautomatiseerde ondersteuning bij het navigeren van drones is zo goed dat er bijna geen training meer nodig is om ermee te vliegen. Het blijft nog wel een uitdaging om het aantal operators te reduceren. Er zijn op dit moment een of meer operators nodig om een drone te besturen wat de aansturing van meer dan één drone nog een uitdaging maakt. Ook de batterijduur legt nog beperkingen op aan wat er met een drone mogelijk is. Het steeds verder gaande gemak waarmee bewapende drones zullen kunnen worden ingezet onderscheidt deze technologie echter wel van bijvoorbeeld chemische wapens waarvoor meer expertise is vereist; in tegenstelling tot chemische wapens kunnen drones voor een zeer specifiek target worden ingezet en op een meer proportionele wijze worden ingezet.

**Killer robots zijn autonome systemen die zonder menselijke tussenkomst op basis van 'rules of engagement' beslissen over leven of dood**. Zulke robots kunnen verschillende vormen hebben: drones, autonome voertuigen, onderzeevaartuigen, etc. De drones die momenteel offensief worden ingezet in b.v. Afghanistan en het Midden Oosten vallen *niet* onder de term "killer robots", want de cruciale beslissing om te doden wordt door een menselijke operator (op afstand) genomen. Uit eerdere paragrafen zal het duidelijk zijn dat techniek en wetenschap voor de afzienbare toekomst niet in staat zijn om de software voor zulke systemen betrouwbaar te construeren[xiv].

**Kunstmatige Intelligentie kan niet alleen offensief worden ingezet maar ook defensief om ons te beschermen tegen calamiteiten**. KI kan bijvoorbeeld worden ingezet om drones te detecteren. KI kan ook worden ingezet om te voorspellen waar de kans op ongewenste situaties zoals bijvoorbeeld een inbraak het grootste is om preventief te kunnen ingrijpen. De strikte regelgeving als het gaat om de toepassing van drones draagt wellicht bij aan de veiligheid maar werpt ook barrières op die het moeilijker maken om deze innovatieve toepassingen in Nederland te ontwikkelen. Het risico bestaat dat Nederland zich op termijn meer afhankelijk maakt van derden om toegang tot geavanceerde technologie te krijgen in plaats van zelf een voortrekkersrol in deze ontwikkelingen te spelen.

(*) Met dank aan Guszti Eiben, Frank van Harmelen, en Mike Ligthart voor feedback op eerdere versies.

---

[i] Jacopo Urbani, Spyros Kotoulas, Jason Maassen, Frank Van Harmelen, Henri Bal (2012). *WebPIE: A Web-scale Parallel Inference Engine using MapReduce*, Journal of Web Semantics, Volume 10, pp. 59-75.

[ii] Yoav Shoham, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Niebles, Terah Lyons, John Etchemendy, Barbara Grosz and Zoe Bauer (2018). *The AI Index 2018 Annual Report*, AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA. Te vinden via: https://aiindex.org/.

[iii] Zie bijvoorbeeld Babak Ehteshami Bejnordi, et al. (2017). *Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer*. JAMA. Vol. 318(22), pp. 2199–2210 en A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, and S. Thrun (2017). *Dermatologist-level classification of skin cancer with deep neural networks*. Nature, Vol. 542(7639), p.115.

[iv] Zie https://blogs.microsoft.com/ai/machine-translation-news-test-set-human-parity/.

[v] "The tasks [that exceed human-level performance] are highly specific, and the achievements, while impressive, say nothing about the ability of the systems to generalize to other tasks." AI Index report, zie voetnoot ii.

[vi] Elsevier 2018, *Artificial Intelligence: How knowledge is created, transferred, and used. Trends in China, Europe, and the United States*. Te vinden via: https://www.elsevier.com/research-intelligence/resource-library/ai-report.

[vii] Zie noot ii.

[viii] AINED 2018, *AI voor Nederland*. Te vinden via: https://www.vno-ncw.nl/nieuws/bedrijven-en-wetenschappers-willen-nationale-ai-strategie.

[ix] SIGAI 2018, *Dutch AI Manifesto*. Te vinden via: http://bnvki.org/.

[x] Veel robotici hebben het over autonome systemen als er geen interventie van mensen nodig is, wat afwijkt van het meer gangbare betekenis van het woord die in de hoofdtekst wordt gebruikt (autonomie gaat over het nemen van geïnformeerde beslissingen).

[xi] Denk bijvoorbeeld aan het verschil in het maken van afwegingen die locatie speelt. Als een marineschip voor de Nederlandse kust of voor de kust van Rusland een vliegend object detecteert, dan heeft dat een andere betekenis op basis van de historische en politieke context.

[xii] Zie ook: Mica R. Endsley, (2017). *From Here to Autonomy: Lessons Learned From Human–Automation Research*. Human Factors, 59(1), pp. 5–27.

[xiii] N. Akhtar and A. Mian, (2018). *Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey*, in IEEE Access, vol. 6, pp. 14410-14430.

[xiv] In een meer systematische analyse zouden we verschillende type fouten kunnen onderscheiden: (a) fouten die te maken hebben met (geautomatiseerde) perceptie, b.v. het niet herkennen van een drone, (b) het gebrek aan adequaat meenemen van context door KI, en (c) een overschatting van mensen van de autonomie van KI.