



MAXIM FEBRUARI

## Goede robotsoldaat wijkt net als mens van regels af

**B**innenkort wordt in onze gemeente het restafval niet meer opgehaald. Alles wat overblijft na het grote afval scheiden moet je zelf achterop de motor naar de milieustraat brengen. Zelf het gemeentelijke groen snoeien. Zelf de burenen in bad doen. Nu de robotisering doorzet, hoeven we ons niet langer het hoofd te breken over politiek, oorlog en financiën: de grote vragen van ons bestaan worden door superieure software voor ons beantwoord. Zelf hoeven we alleen nog het vuilnis op te halen, de straat te vegen en de putjes te scheppen. Een droombestaan.

Toch is het de vraag of we niet zelf ook een oogje in het zeil moeten houden als het gaat om de serieuze kwesties. Om oorlogvoering, bijvoorbeeld. De wereld neigt tot de gedachte dat machtsblokken hun onderlinge vetes voortaan wel kunnen laten uitvechten door tanks en vliegtuigen met autonome systemen aan boord. Dat heeft duidelijk voordelen. De belangrijkste daarvan is dat je in systemen, anders dan in mensen, de Geneefse Conventies kunt inbouwen. Zo voorkom je dat de strijdende partijen het oorlogsrecht schenden.

Maar zelfs als onbemande, autonome strijdkrachten beloven de internationale verdragen netjes te volgen, is het maar helemaal de vraag of het recht is gediend bij dat inbouwen. Wat is recht eigenlijk, en wat behoeft het? De kwestie wordt door roboticisten en robotfilosofen nogal eens laconiek genegeerd.

### *De piloot zegt: vlucht mislukt als je regels gehoorzaam volgt*

**T**en eerste gaan onderzoekers vaak snel voorbij aan het verschil tussen een regel die op papier is vastgelegd en een regel die is verwerkt tot code. Een papieren regel speelt een rol in het eeuwig durende menselijke gesprek over recht en rechtvaardigheid: hij is analoog, niet eenduidig, hij laat ruimte voor interpretatie, hij wordt terzijde geschoven, omgebogen, opgerekt, toegesneden op het geval. Die flexibiliteit verdwijnt ogenblikkelijk zodra je een rechtsregel of morele regel digitaliseert en hem in een systeem stopt.

Nu klinkt die rigiditeit van de gedigitaliseerde regel misschien juist aantrekkelijk. 'Het laatste wat je wilt', zegt een onderzoeker van het Georgia Institute of Technology in Atlanta, 'is een autonome robot op een militaire missie sturen en hem ter plekke laten uitzoeken welke ethische regels hij zal vol-

gen.' Deze Ronald Arkin geeft daarom, net als de meeste van zijn collega's en het leger, de voorkeur aan robots die een affe moraal meekrijgen in hun ransel. Een set van regels die in alle denkbare gevallen leiden tot een vooraf voorziene beslissing.

Helaas. Hier stuit je op het levensfeit dat de wereld fundamenteel onvoorspelbaar is. Piloten heb ik vaak horen zeggen dat elke vlucht onherroepelijk zou mislukken als je alle regels steeds gehoorzaam zou volgen: in iedere vlucht doet zich wel een uitzondering voor die vraagt om aanpassing van de regel aan het geval. En wat geldt voor vliegen, geldt eens te meer voor de ongrijpbare praktijk van oorlogsrecht en moraal. Je kunt niet alles zien aankomen. De betere robotethici begrijpen dat natuurlijk ook wel, en dus maken ze menselijk ingrijpen mogelijk wanneer de autonome systemen dreigen te blunderen. Waarmee je terug bent bij de rol van de mens als interpretator van de verdragen.

Want wie mag dan ingrijpen? En op grond van welke criteria? En wie heeft die programma's überhaupt laten ontwerpen? Als dat de machthebbers zijn, en natuurlijk zijn dat de machthebbers, hoe groot is voor hen dan de verleiding om de Geneefse Conventies al bij het programmeren hun eigen kant op te interpreteren? Retorische vragen. Ik kan beter zeggen: de verleiding is groot om de Geneefse Conventies al bij het programmeren hun eigen kant op te interpreteren.

**L**os van zulke machtsdromen laten moraal en recht zich niet vooraf construeren. Als ik het niettemin interessant vind, de ontwikkeling van juridische en ethische software, dan is het om analytische redenen. De ontwerpers laten zien uit welke elementen een beslissing zoal bestaat: regels, principes, precedenten, emoties, redeneringen, patronen. Hun programma's zijn een nuttige gids bij het denken.

Maar moraal zelf is een praxis. Een handelingspraktijk. Een praktijk waaruit weliswaar regels komen bovendien, als nuttige richtsnoeren voor de toekomst, maar die regels zijn niet het handelen zelf. En zo is ook oorlogsrecht een praxis: de rechtvaardigheid van het handelen vormt zich in het handelen. Daar hoort de vrijheid bij om niet te gehoorzamen aan regels die vooraf zijn meegegeven. En de verantwoordelijkheid om oordelen te trotseren die achteraf worden gevormd. Bouw je een gedigitaliseerde morele of juridische code in, dan maak je ongehoorzaamheid onmogelijk en neem je verantwoordelijkheid weg. Zo raken mensen moreel lam en machteloos. Ze sneuvelen alleen nog wel zelf.