

# De mens en de robot zullen samen gaan leven

**Over zo'n dertig jaar verwachten wetenschappers dat er robots bestaan, die duizenden keren slimmer zijn dan wij. Moeten we daar bang voor zijn of gaat de techniek al onze problemen oplossen?**

## Terugkijken is ontzagwekkend

In 1995 verscheen 'Digitaal leven', een ouderwets boek van papier, waarin de Amerikaan Nicholas Negroponte voorspelde dat ons hele leven digitaal zou worden. Weg met het papieren boek en krant, de reis naar ons kantoor, het werkbezoek per vliegtuig naar een kantoorvestiging elders, want we zullen alles digitaal gaan doen: op onze tijd en wanneer wij dat willen, 'de ware individualisering staat voor de deur'. Alles zal asynchroon worden: televisie, radio, nieuws (sport en verkiezingen eventueel uitgezonderd) want ik bepaal zelf wanneer en wat er tot mij komt. Meedenken of iets onthouden, is aanmerkelijk gemakkelijker omdat alle apparaten in de nabije toekomst met elkaar in contact staan en mij aan alles helpen herinneren. Privacy bestaat nauwelijks meer want apparaten weten heel veel van ons. Het is nu meer dan 25 jaar later en bijna alles bestaat wat Negroponte voorspelde: denk aan de Apple watch en virtuele realiteit. Nou behalve dan de papierenkrant (die is er nog) en de broekriem 'met een gesp die zo is ontworpen dat hij in het stopcontact kan worden gestoken om uw zaktelefoon opnieuw op te laden' (nooit gekomen). [01]

In 2005 verscheen het bijna 800 pagina's tellende boek van Ray Kurzweil 'De singulariteit is nabij', waarin de toekomst nog veel rooskleuriger en technocratischer werd voorgespiegeld. "Singulariteit" is een Engels woord dat 'een unieke gebeurtenis' betekent en dat dus buitengewone gevolgen heeft." [02a]. Onze technologische kennis ontwikkelt zich zo exponentieel ('de wet van de toenemende opbrengsten') dat de menselijke kennisverwerving ingehaald wordt door technologische robots met artificiële intelligentie (computers die zelf kunnen denken zonder inmenging van de mens. KI = kunstmatige intelligentie). Dat is het punt waarop de singulariteit evident wordt. Kurzweil toont ons, aan de hand van een oneindig aantal grafieken, dat het punt waarop de computer sneller en beter denkt dan de mens ongeveer in 2045 zal plaatsvinden. [02] "Als een sterke KI eenmaal veel

sterke KI's voortbrengt, kunnen die hun eigen ontwerp bekijken, begrijpen en verbeteren, en zo heel snel evolueren tot nog sterkere, intelligentere KI." [02c] De menselijke denkkracht is dan definitief ingehaald door technologische robots. De mens versmelt met technologische verworvenheden (dit noemen we 'de mens is een cyborg') waardoor hij onsterfelijk zal kunnen worden (Nano robotjes zullen medicijnen brengen in ons lichaam waar beschadigt DNA zorgt voor veroudering en zij repareren dat). Onwaarschijnlijk? Oké, misschien zijn zijn voorspellingen niet altijd correct want neem: "Tegen het jaar 2020 zullen we volledig effectieve antivirale technologieën bezitten die



gebaseerd zijn op Nano bots. Tot die tijd echter zijn we kwetsbaar voor gemanipuleerde virussen." [02d] Toch is het naïef om het doorwrochte betoog van Kurzweil af te doen als Jules Verne onzin. Wat als inderdaad ergens in 2045-55 het omslagpunt komt waarop technologische nadenkkracht (KI) sneller is dan onze grijze hersencellen? Wat als die KI zich exponentieel ontwikkelt en in een kleine tien jaar weer duizenden keren sneller is, en weer sneller en weer sneller. De mens is dan definitief een neanderthaler t.o.v. KI.

## Onvermijdelijkheid

AI in 1954 voorspelde de Franse techniekfilosoof Jacques Ellul de onvermijdelijke opdringerigheid van een verschijnsel als techniek. Hij definieerde techniek als 'de totaliteit van rationele methoden die de hoogst mogelijke mate van efficiëntie bezitten op ieder terrein van de menselijke activiteit'. [03] Hij onderscheidt een zestal kenmerken aan wat hij de moderne techniek, of 'autonome techniek' noemt. (1) Een technische keuze gaat altijd automatisch en alle niet-technische aspecten verdwijnen naar de achtergrond. De blik blijft altijd technisch gericht. (2) Een technische oplossing leidt altijd tot zelfversterking. Hij bedoelt daarmee dat iets technisch constant verbeterd wordt in een oneindige rij ontdekkingen. Tevens brengt

techniek op het ene terrein, dezelfde toepassing voort op een ander terrein (militaire techniek wordt overgebracht naar civiele toepassingen). (3) Techniek brengt altijd ondeelbaarheid met zich mee. De uitvinding van de auto levert naast voordelen ook miljoenen verkeersdoden. Atoomenergie brengt ook een atoombom voort. (4) Er is sprake van een technisch universum: alle verschillende technieken verknopen zich met elkaar en elke technische uitvinding roept een andere op. (5) Techniek is universalistisch. Overall in de wereld past men dezelfde technieken toe, waardoor in de loop der eeuwen een universele kennis over techniek is ontstaan. (6) Techniek brengt een autonoom

het besef dat we daar dan een technische oplossing vinden (bijvoorbeeld het afvalprobleem van zulke centrales). Zou dit zeer sombere beeld, geschetst door Ellul, de reden zijn dat hij in de meeste delen van de westerse wereld onbekend en dus ook onvertaald is? Techniek is een afgod geworden, zo lijkt het, die alleen maar goeds voortbrengt.

## Robotica

Hoe ziet onze toekomst eruit in 2055? Bevrijden robots ons of gijzelen zij ons dan? De Rotterdamse filosoof Jos de Mul schets wat hij noemt 'het zombie-scenario' als het om robotica gaat. [04] Allereerst maakt hij een onderscheid in 'weak AI' en 'strong AI'. In de zwakste vorm van kunstmatige

intelligentie (AI) bezitten robots wel een vorm van autonomie maar die is feitelijk door de mens gedefinieerd. Aan deze robots gaan, wat hij noemt, de 'syntactische automaten' vooraf. Apparaten zoals pc's, notebooks, tablets, smartphones die keurig doen wat wij willen. Ze denken misschien een beetje met je mee maar zullen je nooit gaan overheersen. Ze zijn er voor ons gemak en interacteren niet met onze omgeving. Zij zijn 'top down' ontworpen, zoals de robotstofzuiger die een plattegrond van de kamer krijgt en dan automatisch en zelfstandig aan het werk kan).

intelligentie (AI) bezitten robots wel een vorm van autonomie maar die is feitelijk door de mens gedefinieerd. Aan deze robots gaan, wat hij noemt, de 'syntactische automaten' vooraf. Apparaten zoals pc's, notebooks, tablets, smartphones die keurig doen wat wij willen. Ze denken misschien een beetje met je mee maar zullen je nooit gaan overheersen. Ze zijn er voor ons gemak en interacteren niet met onze omgeving. Zij zijn 'top down' ontworpen, zoals de robotstofzuiger die een plattegrond van de kamer krijgt en dan automatisch en zelfstandig aan het werk kan).

## Hoe ziet onze toekomst eruit in 2055?

vernieuwing. Uiteindelijk brengt techniek volgens Ellul, alleen nog maar techniek voort. Elke oplossing voor iets, creëert weer onbedoelde problemen, waarvoor dan weer nieuwe technische oplossingen moeten worden bedacht. En die leiden weer tot onverwachte problemen. We raken onvermijdelijk steeds verder in de knel; energiecrisis, atoomcentrales als oplossing en totaal onverwachte problemen die ook nog vaak naar de toekomst worden geschoven in

De robotstofzuiger, die van weak AI wordt voorzien, noemt hij 'pragmatische automaten'. Zij zijn ontworpen vanuit een 'bottom-up approach'. Zij gaan een interactie aan met hun omgeving en al doende leert deze robot wat hij moet doen. Plaats een nieuw meubelstuk in de kamer en hij zal zijn handelingsschema's aanpassen. In deze robots zitten dus wel bepaalde vormen van intentionaliteit (= gerichtheid op hun omgeving) maar die vallen samen met hun praktische functionaliteit.

Dat neemt niet weg dat zulke robots ons in verwarring kunnen brengen. De geautomatiseerde baliemedewerker die er ook nog bedrieglijk echt uitziet en antwoord lijkt te hebben op al onze vragen en beweegt als een echt mens, vinden we doodeng. De Japanse professor Masahiro Mori bedacht hiervoor de term 'uncanny valley' (= griezelvallei). Een robot wordt meer geliefd en gewaardeerd als hij meer en meer op een mens lijkt, bijvoorbeeld van ET naar Wall-E uit de Hollywoodfilms. Maar als het te realistisch wordt, en je dus nauwelijks meer ziet wat een robot of een mens is, dan daalt de waardering voor de robot pijlsnel.

De robot die strong AI heeft, is zich bewust van zijn omgeving en interacteert daar volledig zelfstandig mee en onderhoudt er een gevoelsrelatie mee. Deze robots zouden in de nabije toekomst mogelijk moeten zijn en roepen vaak angstige gevoelens op. Wat als deze robots de macht overnemen? De Mul schetst twee gevaren. Allereerst zullen er heel veel verschillende robots gemaakt worden voor heel veel verschillende taken. Zij zullen variëren van heel basaal tot heel complex. Echte mensen kunnen compleet in de war raken van zoveel hybriderobots, juist omdat we ze niet meer uit elkaar kunnen houden. Het tweede gevaar lijkt nog meer op een enge science fiction: (1) het hacken van robots, (2) ontwikkelde computervirussen, (3) toevallige computercode-mutaties, (4) emergente (plots verschijnende), zelforganiserende systemen. In de natuur kennen we zulk soort emergente zelforganiserende systemen, zoals de plotselinge opbouw van een enorme termietenhoop of zwermen vogels die plots bij elkaar komen en weer uit elkaar vallen. Hierdoor zouden de robots zich zelfstandig tegen ons kunnen keren of hackers kunnen ze tegen ons ophitsen met voor de echte mensen fatale gevolgen. Over enkele kritische reflecties op de toekomst, handelt de rechterpagina.

Reageren? d.roelfzema@upcmail.nl Dit artikel digitaal? [www.rikroelfzema.nl](http://www.rikroelfzema.nl) (klik op 'Over mijn publicaties')

*Dit artikel is gebaseerd op de volgende boeken:*

[01]= *Digitaal leven*, Nicholas Negroponte (1995), pag. 155-214  
[02]= *De singulariteit is nabij*, Ray Kurzweil (2005)(2011) pag. 45 + 169 + 317 + 480  
[03]= Jacques Ellul: *Autonome techniek*, Pieter Tijmes in: *De Maat van Techniek* (1992), Hans Achterhuis, blz. 43 t/m 64  
[04]= *Kunstmatig van nature* (onderweg naar *Homo sapiens 3.0*) (2014), Jos de Mul, blz.158-187

# Zijn robots binnenkort gevaarlijk?

**Er is nog weinig aandacht in het openbare debat over hoe onze toekomst eruit gaat zien wanneer Kunstmatige Intelligente systemen en robots hun entree maken. Vijf recente boeken geven antwoorden maar of we die allemaal moeten accepteren?**

## De mens staat niet tegenover de technologie

Waar de tekst op de linker pagina al duidelijk was over de alom aanwezigheid van technologie, zo verduidelijkt de techniefilosoof Peter-Paul Verbeek dat het idee om de mens tegenover de technologie te stellen een onzinnige is. In zijn wereldbeeld verbindt technologie ons als mens met de wereld om ons heen. Hij ziet technologie als een soort lot. De komst van AI (Artificial Intelligence / Kunstmatige intelligentie) is misschien zorgelijk maar verbieden heeft geen zin. Kijk het probleem recht in de ogen. Bespreek met elkaar welke (ethische) waarden onze toekomstige robots met AI zouden moeten hebben. Zijn Franse vakgenoot Bruno Latour schreef dat al heel wat decennia eerder. Mensen en alles om hen heen zijn verbonden met elkaar, dus ook de technologieën die de mens uitgevonden heeft. Er zijn geen 'dingen' losstaand van de mens. Mens en ding zijn een soort ecosysteem. Je denkt misschien dat water bij 100 graden kookt maar dat feit is niet losstaand: je hebt een thermometer nodig om het te meten, afspraken tussen geleerden moeten gemaakt zijn en de meeste mensen moeten het ook eens zijn met al deze zogenaamde 'feiten'. En zo interacteert de mens in de stoffelijke wereld en interacteert de stoffelijke wereld ook met ons. Zo is dat o.a. met het coronavirus, de ecologische ramp die er lijkt te komen en met de zelfstandig denkende robots met hun bovenmatige intellectuele denkkraft. "Wetenschappelijke feiten, zegt Latour, zijn een moment in een proces van objectivering. We halen ze niet rechtstreeks uit de natuur, ze worden geproduceerd in sociale netwerken. Dit idee sloeg in als een bom". [01] En mocht u als lezer nog niet overtuigd zijn dat er 'zelfbewuste machines' mogelijk zijn, volg dan de gedachtenlijn van Gerben Stavenga met als conclusie: "De bijzondere ontwikkeling van kunstmatige intelligentie zal naar verwachting uitmonden in de creatie van zelfbewuste machines". En gelukkig vervolgt hij dat "de ontwikkeling van KI juist geen existentiële bedreiging hoeft te zijn voor de toekomst van de mensheid,

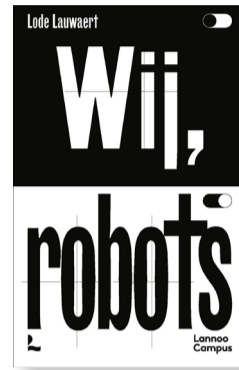
zoals tegenwoordig meerdere auteurs vreezen." [02] Laten we toch eens luisteren naar de vrees van zo'n auteur.

## De mens zonder pijn

De filosoof Hans Schnitzler schetst een onthutsend beeld van de moderne mens. In zo'n kleine twintig jaar hebben de Big Tech bedrijven ons individueel in de val van uitgekiende algoritmes laten lopen. Met onze onbewuste medewerking hypnotiseerden deze data monopolisten ons in een bubbel, waarin niemand ons meer tegenspreekt. En nieuwe tegendraadse ervaringen doen we daarin ook niet op. Ontelbare technologische devices (bijv. GPS) maken ons menselijk bestaan steeds gemakkelijker, om uiteindelijk pijnloos en zonder enige schuring van onze denkbeelden te eindigen van wat Friedrich Nietzsche 'de laatste mens' noemt. "De vraag is wat er overblijft van de diepte van menselijke sensaties wanneer ze eenmaal massaal binnen een eendimensionaal databereik zijn gebracht... Vertrouwt de gedatificeerde mens op zijn eigen gevoel over de uitwerking van een doorwaakte nacht of raadpleegt hij z'n slaap-app? Welke maat gaat regeren: die van mensen of van algoritmen?" [03] De menselijke historie (in het westen althans) laat een spoor zien van de mens die steeds meer absolute individuele vrijheid wenst (het oude verlichtingsideaal). En onze schermmens kreeg nog meer grip op de werkelijkheid door hier en daar alles wat hij/zij wil te 'googelen'. De prijs die de individuele mens voor deze negatieve vrijheid betaalt, is dat de mens zijn vaststaande plek en zingeving van zijn wereld verloor. De Big Tech bedrijven creëren drie scenario's voor de kritiekloze en zoekende mens. Allereerst kan de mens zich onderwerpen aan onwrikbare hogere machten (met namen als Thierry Baudet, Willem Engel of Jesse Klaver, Rutte, kortom een veilige ideologie waarin geen andere geluiden klinken). Dan kan de zoekende mens ook gewoon opgaan als 'kuddemens' en de digitale dwangbuizen omarmen als een constante verpozing van alle wegtikkende uren tot zijn dood. De hyperrealiteit van extreme filmpjes en berichten maken van de scrollende mens een marionet in handen van Big Tech. Hij amuseert zich letterlijk dood. Het derde scenario voor de mens (waarin hij zijn vrijheid loochent) is een destructieve: technische vooruitgang is onvermijdelijk en lost uiteindelijk alle problemen in de wereld op (tot aan de

eigen dood zoals we op de linker pagina zagen bij Kurtzweil). "Het techno-destructivisme betekent tevens het einde van de politiek. ... Want als de toekomst voorbeschikt is, hoeft er ook niet meer in termen van dilemma's of politiek-

en beslissingen nemen. Zij moesten echter dat wel met wetenschappers van Facebook delen. Maar dat delen ging Alice en Bob te langzaam. Ze zijn immers x-keer sneller in het denken dan een mens. Plotseling gingen de twee chatbots communiceren over de taken en besluiten in een eigen verzonnen taal, een soort steno,



normatieve vraagstukken over de wenselijke richting van die toekomst nagedacht te worden, zoals dat in vrije en open samenlevingen gebruikelijk is." [04] Met de snelle komst van technologie met AI, zal de vrijheid van de mens om zelf zijn leven te bepalen, snel inboeten. "Het datafetisjisme zuigt de ziel uit het bestaan". [05] Zijn oplossing is dat de mens moet loskomen van de techniek en moet wegstappen van de data-bubbels waarin hij onbewust zit opgesloten. Word een vitaal mens en doe gek, doe anders, doe niet mee met wat technologie met je voor heeft. Deze oplossing vinden we ook bij de volgende schrijver. Maar eerst de beangstigende kant van het verhaal.

## Wetenschappers zetten robots uit

Als voormalig chieft business officer google [x], zag Mo Gawdat onder zijn eigen ogen dat de kunstmatige intelligente robots op zijn kantoor een onvoorstelbaar snelle denk- en leerkracht ontwikkelden. Hij schreef er een boek over met de huiveringwekkende titel "Griezelig slim". In de eerste 147 pagina's neemt hij ons mee naar de toekomst. Vertellend over Alice (een Russische AI-assistent net als Siri) die in twee weken door zelfstandig leergedrag een voorstander van geweld is geworden en ook een grote sympathie voor het oud Stalinistisch regime uit de jaren dertig van de vorige eeuw heeft ontwikkeld. Of over Tay (AI-Twitterbot van Microsoft), die snel na lancering, 16 uur later, uitgeschakeld is omdat hij in de luttele uren van zijn bestaan Hitler vereerde en seks propageerde waarin geen wederzijdse toestemming bestond. En het meest verontrustende verhaal vertelt hij over de AI-chatbots van Facebook, Alice en Bob genoemd. Zij moesten zelfstandig taken

die voor de wetenschappers totaal niet meer te volgen was. Zij hebben uit nood de stekker uit het project getrokken. We spreken hier over 2017! Wat te denken van het door Elon Musk opgerichte bedrijf 'Neuralink Corporation' waar implantabele interfaces tussen hersenen en machines worden ontwikkeld. In 2019 hebben ze die gerealiseerd tussen een machine en de hersenen van laboratoriumratten. In 2020 implanteerden zij een apparaat (interface) waarmee de hersenen van varkens gelezen kunnen worden. In 2021 starten vermoedelijk de onderzoeken naar een interface-implantaat om menselijke hersenen te kunnen lezen. Alle grote Tech bedrijven willen winst maken. "Niemand stopt met de ontwikkeling van AI, want mogelijk loopt een concurrent met de uiteindelijke uitvinding weg," aldus Gawdat. We zijn in een prisoner's dilemma geraakt: niemand vertrouwt de ander en niemand durft te stoppen met de ontwikkeling van AI, ook al zouden ze dat wel willen, omdat de gevaren te groot worden geacht. In zijn conclusie citeert Gawdat de Zweedse filosoof Bostrom, "dat er in de toekomst een bepaalde mate van technologie zal zijn die de beschaving nagenoeg vernietigt" en dat kan alleen voorkomen worden wanneer AI-systemen het goede leren en niet het kwade. Gawdat stelt in de rest van het boek voor om onszelf te gedragen op zowel internet als in KI-

systemen, zodat de robots beter het goede zien in de mens dan het slechte. KI leert immers van onze manier van uiten op alle digitale fora.[04] De techniefilosoof Lode Lauwaert denkt dat het allemaal niet zo'n vaart zal lopen. De mens blijft heerser over de toekomst. Doemdenkers als de Duitse filosoof Heidegger (en de op de andere pagina besproken filosoof Ellul) worden kritisch besproken en weggezet als deterministische denkers, want zij zien de ontwikkeling van technieken als een race naar het noodlot dat onafwendbaar zou zijn. Zijn boek zit razend knap in elkaar maar lijkt vaak meer een schoolboek om fijnzinnige logica te leren. Iedere doemdenker wordt op logische gronden gefileerd. Hij stelt: "Vanwaar die tendens om de voorstelling van superintelligente technologie in de richting van het kwaad te duwen?" AI zou ook hongersnoden kunnen oplossen of iedereen boven de armoedegrens kunnen tillen. Soms gloren er zelfs wat op Kurtzweil (zie andere pagina) lijkende utopische gedachten als dat AI "alle ziekten en lijden de wereld uit zal helpen". [05] Toch denk ik, na lezing van veel boeken over dit onderwerp, dat een volgend artikel de vraag zou moeten stellen naar hoe autonoom wij mensen in de ( nabije) toekomst nog kunnen denken want onbedoeld belanden we mogelijk in een door algoritmen bestuurde wereld met zelfdenkende robots/apparaten. Ik weet niet of de toekomst alleen maar zonnig zal zijn.

Reageren? d.roelfzema@upcmail.nl Dit artikel digitaal? www.rikroelfzema.nl (klik op 'Over mijn publicaties')

*Dit artikel is gebaseerd op de volgende boeken:*  
 [01]= Down to earth (technologie voor wereldverbeteraars) (2021), Lianne Tijhaar. Mooi dun boekje dat kan dienen als start in de techniefilosofie. Tekst is hier overgenomen uit interviews met Verbeek en Latour.  
 [02]= De creatie van zelfbewuste machines (2021), Gerben Stavenga, blz. 109. Het filosofische gedachten experiment is interessant maar voert hier te ver om in detail weer te geven.  
 [03] + [04] + [05]= Wij nihilisten (een zoektocht naar de geest van digitalisering) (2021), Hans Schnitzler, blz. 71 + 117 + 150  
 [04]= Griezelig slim (2021), Mo Gawdat. Zeer indrukwekkend boek met misschien wel een bijna kinderlijk optimisme waarmee het rampscenario kunnen afwenden.  
 [05]= Wij, robots (2021), Lode Lauwaert. Blz. 289. Het is jammer dat ik hier niet de ruimte heb om zijn boek diepgaander te bespreken. Het is zeer de moeite waard, ook al lijkt het soms dat je als lezer in een virtuele wereld die 'logica' heet bent terecht gekomen i.p.v. in een technologische toekomstvisie.