

## Ondenkbare machine

### Hernieuwd debat over kunstmatige intelligentie

*De Academische Boekengids 37, februari 2003, pp. 4-6.*

**De droom van de denkende computer uit de jaren zestig is terug. Toch komt het onderzoek wellicht nooit uit het domein van de verbeelding. Drie boeken stemmen optimistisch. Eén zet het vraagteken.**

Aan het begin van de eenentwintigste eeuw lijkt de kunstmatige intelligentie bezig te zijn met een opmerkelijke comeback. Kenmerkend voor de hernieuwde belangstelling in kunstmatige intelligentie, ook wel *artificial intelligence* of kortweg 'AI' genoemd, is dat opnieuw alles draait om systemen die menselijk of dierlijk gedrag nadoen, met name het soort gedrag waar logisch denken en redeneren voor nodig is. Maar dit keer is er iets nieuws bij gekomen: het internet. De mogelijkheden van het wereldwijde web zijn gigantisch, omdat in principe de rekenkracht van alle computers op onze aardbol met elkaar kan worden verbonden. De vraag is of het in dit nieuwe millennium eindelijk zal lukken om met behulp van het alwetende netwerk het intelligente wezen, de denkende robot, de slimme machine te vervaardigen. Is de techniek eindelijk zover?

*De afkorting AI werd in de jaren negentig dikwijls met een zure glimlach gelezen als almost implemented.*

De verwachtingen zijn opnieuw hooggespannen, ook buiten de wetenschap. Er verschijnen weer films in de geest van Stanley Kubricks klassieker '2001: A Space Odyssey', waarin robots niet alleen sprekend op mensen lijken, maar ook nog kunnen denken en voelen als mensen, zoals ook in Steven Spielbergs film 'A.I.' In laboratoria wordt weer meer AI-onderzoek verricht, en er verschijnen elektronische producten op de markt die expliciet verwijzen naar de hoogtijdagen van de *artificial intelligence*, zoals Sony's robothondje Aibo. De naam van het hondje is een verbastering van 'AI' en 'bot', wat samen in het Japans 'vriendje' betekent, en wat ook goed aangeeft dat de 'intelligentie' van Aibo is gebaseerd op programma's en lessen die uit de oude vertrouwde AI-doos stammen. Zelfs Bill Gates heeft AI weer als vriend binnengehaald: hij beloofde onlangs op de 'International Joint Conference on Artificial Intelligence' een doorbraak in de kunstmatige intelligentie, waarmee gebruikers mobiel door slimme computers ondersteund worden.

Bij zoveel belangstelling voor een opwindend thema, waarbij de techniek onophoudelijk en in een razend tempo voortgang boekt, blijven uitgevers niet achter. En dus verschijnen momenteel aan de lopende band non-fictieboeken waarin AI een hoofdrol speelt. De vier boeken die hier worden voorgesteld zijn ieder exemplarisch voor een kenmerkend standpunt in de hernieuwde belangstelling voor AI.

De droom van denkende machines begon voor het eerst concrete gestalte aan te nemen op een conferentie aan het Dartmouth College in Hanover/New Hampshire (VS). We schrijven 1956. Tientallen vooraanstaande wetenschappers, waaronder de wiskundigen Claude Shannon en John McCarty en de neuroloog-wiskundige Marvin Minsky, bedachten dat *artificial intelligence* een sexy naam was voor een veelbelovend onderzoeksprogramma waarin zulke uiteenlopende vakgebieden als psychologie, neurologie, biologie, elektrotechniek, linguïstiek, wiskunde, informatica en logica hun krachten zouden gaan bundelen. De algemene verwachting was destijds dat de realisatie van digitale computers of computergestuurde robots, in staat om een vorm van hogere intellectuele processen te simuleren, binnen handbereik lag. Minsky verwoordde de algemene euforie nog het mooist. 'Within a generation', bulderde hij, 'the problem of creating "artificial intelligence" will be substantially solved.'

Weer tien jaar later, door de komst van de Personal Computer, kwamen de verwachtingen rond AI tot een hoogtepunt. Er verschenen prototypes van medische-diagnoseprogramma's en grootschalige onderzoeksprogramma's naar software voor spraakherkenning. Vanuit allerlei technische en wetenschappelijke hoeken waren optimistische geluiden te horen. Op een conferentie van de American Association of Artificial Intelligence (AAAI), gehouden in 1985, verwachtten duizenden onderzoekers het nieuws van de beslissende doorbraak te horen.

Maar het nieuws bleef uit. Wat een 'Blitzkrieg' had moeten zijn, werd een slopende veldslag. De droom van het machinaal simuleren van natuurlijke intelligentie bleek niet zo makkelijk te realiseren als oorspronkelijk gedacht. Overall rezen lastige praktische beperkingen, zowel in de hardware (de chips en bedradingen) als in de software (de instructies die de chips besturen). Ook veel theoretische problemen bleken helemaal niet zo makkelijk op te lossen als aanvankelijk gedacht. Terwijl de ene generatie Personal Computers na de andere in hoog tempo van de lopende band rolde, was van de Denkende Computer nog geen spoor te bekennen. En de machinale vertalingen, de computermodellen die de beurskoersen trefzeker konden voorspellen, en de tientallen 'expert-systemen' bleven ofwel steken in een infantiel stadium, ofwel gingen in de praktijk dramatisch onderuit. De in AI-kringen nogal eens gehoorde

voorspelling dat de mens na het jaar 2000 alleen nog maar een soort gemankeerde robot zou zijn, een kreunende zak met genen die het met zijn miserabele reken capaciteit op alle fronten zou afleggen tegen de zacht zoemende dozen vol met razendsnelle Silicium-chips, leek met het naderen van de millenniumwisseling steeds onzinniger.

Het aanvankelijke optimisme was gevoed door de gedachte dat programma's die hun waarde hadden aangetoond in een beperkte, exact gedefinieerde omgeving - zoals die van een schaakbord of een virtuele wereld gecreëerd door wiskundige objecten - 'alleen nog maar' moesten worden aangepast voor een meer algemene context. De impliciete aanname was dat voor die laatste stap geen kwalitatieve, maar alleen kwantitatieve stappen nodig waren. En omdat de techniek onophoudelijk en in een razend tempo voortgang boekt, kon daar het probleem niet liggen. De wet van Moore beschrijft immers dat chips ongeveer iedere anderhalf jaar met het dubbele aantal transistors worden uitgerust, waardoor ze nog meer rekenkracht en nog meer geheugencapaciteit krijgen. En op basis van schattingen was het creëren van intelligentie inderdaad alleen nog maar een kwestie van tijd. Maar deze extrapolaties op het kwantitatieve vlak gingen in kwalitatieve zin allemaal onderuit. 'Bijna' werd een understatement voor 'nooit'. Dat maakte het onderzoek naar AI een klassiek voorbeeld van wat de wetenschapsfilosoof Imre Lakatos eens een 'degenererend onderzoeksprogramma' noemde: het begint allemaal met grote verwachtingen, waarin een nieuwe aanpak in een beperkt gebied indrukwekkende resultaten oplevert. Dan gaan onderzoekers die resultaten in een bredere context in praktijk brengen, en plotseling zijn ze veel minder succesvol. Iedere poging om de resultaten in een bredere context toe te passen, zorgt voor meer bewolking, totdat het uiteindelijk begint te regenen en het hele onderzoeksgebied vast komt te zitten in de modder. De afkorting 'AI' werd in de jaren negentig dikwijls met een wat zure glimlach gelezen als: 'almost implemented'.

Maar tegenwoordig is er weer hoop. Die hoop is niet alleen gevoed door de mogelijkheden van een wereldwijd computernetwerk, maar ook door het nu voor het eerst op de markt komen van concreet bruikbare producten waarin onderdelen van het AI-onderzoek zijn verwerkt. Het hondje Aibo is het bekendste voorbeeld. Ook verschijnt op allerlei gebied programmatuur waarin AI een hoofdrol speelt. Deze programma's worden bijvoorbeeld gebruikt om patronen in bewegende beelden te herkennen, zoals gezichten van voorbijgangers, nummerplaten op langsrazende auto's, of tanks in het struikgewas. Voor dit soort taken is inderdaad een vorm van 'intelligentie' nodig: zo moet een computer voor het herkennen van een gezicht de actuele beelden vergelijken met foto's die in het geheugen zijn opgeslagen, storende omgevingsfactoren uitfilteren, en gezichten kunnen selecteren, ook als die van opzij of bij schemerlicht voor de camera komen.

Al dit soort toepassingen zijn nu pas voor het eerst praktisch bruikbaar. En daarmee keren ook de discussies over de verhouding tussen machinale en menselijke intelligentie weer terug. Onlangs is bij de Duitse uitgeverij Vittorio Klostermann een grondig overzicht verschenen, waarin het jarenlang bakkeleien over kunstmatige intelligentie systematisch op een rij wordt gezet. In zijn boek *Maschine, Körper, Geist* vraagt de cognitief wetenschapper Max Urchs (verbonden aan de Nikolaus Kopernikus Universiteit in Torun) waar we ook alweer waren blijven steken en hoe het er bij de speurtocht naar de mechanismen van intelligentie inmiddels voorstaat. Het eerste deel van zijn werk gaat uitvoerig in op de 'klassiekers' van de cognitieve wetenschappen: zo belicht hij van alle kanten de argumenten voor en tegen de Turingtest - een door de wiskundige Allan M. Turing bedachte proefopstelling waarbij een proefpersoon via een toetsenbord de dialoog aangaat met een voor hem onzichtbaar blijvende menselijke persoon en een machine, om zo te achterhalen of er een verschil tussen beide is waar te nemen. Bij het passeren van alle bekende namen uit het vakgebied, valt op dat Urchs goed het voordeel van de historische afstand gebruikt: elk argument wordt zorgvuldig gewogen zonder daarbij direct in de hitte van de strijd te belanden. Een bekend argument van John Searle tegen de mogelijkheid van AI luidt bijvoorbeeld dat een simulatie van denken nog geen denken is, net zoals de simulatie van het weer nog geen regenbui is, en het simuleren van de spijsvertering geen kruimel verteert. Urchs behoudt de juiste afstand en vraagt zich terecht af: 'Is elke simulatie zo duidelijk te onderscheiden van het gesimuleerde? Kan men zingen simuleren? Of een wereldrecord hordelopen?'

In het tweede deel van het boek komen de actuele voortgangen in het vakgebied aan de orde: wat weten we na jarenlang hersenonderzoek over het functioneren van ons brein? En hoever zijn we gekomen in de computertechniek? Daarbij slaat Urchs' weegschaal allengs door in het voordeel van AI: het tempo waarin het moderne hersenonderzoek bezig is de mechanismen van ons biologisch reken centrum bloot te leggen, de mogelijkheid om die mechanismen in neurale netwerken af te beelden, het internet en de ontwikkelingen in de computerhardware en de robotica maken het voor Urchs waarschijnlijk dat binnenkort een doorbraak mogelijk is in de versmelting tussen menselijke en machinale intelligentie.

Computersystemen kunnen inmiddels al zo klein worden gemaakt, dat er aan te denken valt ze bij mensen te implanteren. De combinatie bio-elektronica en nanotechnologie, de techniek die werkt met de bouwstenen van de natuur - atomen - zou daarbij wel eens voor grote doorbraken kunnen zorgen, aldus Urchs. Daarbij komen mens en machine niet tegenover elkaar te staan, maar naast elkaar. Voor Urchs is zo'n symbiose niet meer dan vanzelfsprekend, omdat de mens al eeuwenlang techniek als uitbreiding en vervanging van zijn ledematen gebruikt. Zolang de blik onbevangen blijft en gericht is op het overschrijden van disciplines en andere grenzen, komt de machinale intelligentie vanzelf.

Vergelijkbaar positieve geluiden komen uit de Verenigde Staten. Rodney Brooks, de auteur van het pas

verschenen en meteen in het Nederlands vertaalde boek *Flesh and Machines. How Robots will change us*, knutselde als kind al robots in elkaar. Hij kreeg bekendheid met zijn 'domme' robots - robots die relatief ingewikkelde taken oplossen door zich op de wezenlijke mechanismen daarvan te concentreren. Rodney is inmiddels directeur van het Artificial Intelligence Laboratory aan het Massachusetts Institute of Technology (MIT). In zijn ronkende boek scheurt de optimist in hoge versnelling langs de succesverhalen uit de korte geschiedenis van de AI. Net als zijn voorganger Marvin Minsky, medeoprichter van het MIT Artificial Intelligence Laboratory, laat Brooks zich verleiden tot krachtige uitspraken als: 'My thesis is that in just twenty years the boundary between fantasy and reality will be rent asunder. Just five years from now that boundary will be breached in ways that are as unimaginable to most people today as daily use of the World Wide Web was ten years ago.'

Juist dit *world wide web* speelt in de toekomstverwachtingen van Brooks een centrale rol. Voor al die taken die met de rekenkracht van compacte computers nog onoplosbaar zijn, zoals het strijken van een zakdoek of het vinden van een stopcontact, kunnen we robots bouwen die via het internet direct verbonden zijn met een veel krachtiger rekencentrum. Blijkt er dan nog meer onoplosbaars, zoals het vangen van een mug of het vragen naar de weg, dan kan een menselijke operator, ergens in de binnenlanden van India of in een ander lagelonenland, bijspringen. Op die manier kun je als leefbaar land zoveel gastarbeiders binnenhalen als je wilt, zonder dat het eigen volk onrustig hoeft te worden. Bovendien valt daarmee het probleem te verdoezelen dat computers 'in het wild', geconfronteerd met de veelzijdigheid van de realiteit, nog niet zo bijster veel kunnen beginnen. Zo zal een alleen met kunstmatige intelligentie uitgeruste huisrobot zonder menselijk ingrijpen geen onderscheid weten te maken tussen bonafide en malafide bezoekers aan de deur.

Brooks' eigen robotbedrijf levert al een robot met internetaansluiting. De iRobot wordt bestuurd (Brooks spreekt van het 'belichamen' van de robot) vanuit een willekeurige internetbrowser. Uiteindelijk, zo verwacht Brooks net als Urchs, zullen mens en machine volledig met elkaar versmelten.

Dat is in grote lijnen ook wat de auteurs beschrijven in het enkele maanden terug verschenen boek *Understanding Artificial Intelligence*, waaronder zo'n beetje alle bekende namen uit het vakgebied verzameld zijn, zoals Marvin Minsky, Ray Kurzweil (bekend van het cultboek *The age of spiritual machines*) en, opnieuw, Rodney Brooks. Allen geven schoorvoetend toe dat de verwachtingen van AI jarenlang wat te hoog gespannen waren, maar dat neemt volgens de auteurs niet weg dat er grote vooruitgang is geboekt.

Een kritische - en tevens frivole - noot in dit gezelschap van heren, komt van een vrouw. De eenendertigjarige Gaby Wood studeerde Frans in Cambridge en werkt nu als journaliste bij *The Observer*. Ze kwam op het idee om Brooks eens een simpele vraag voor te leggen: kijkend naar de geschiedenis van intelligente machines kom je niet alleen in de twintigste eeuw, maar ook in de eeuwen daarvoor al voorbeelden tegen van mensen die geloofden dat het binnen handbereik lag. Sterker, elke eeuw kent zijn eigen Rodney Brooks. In haar boek *Living Dolls. A magical history of the quest for mechanical life* geeft Wood daar voorbeelden van. Zo bedacht in 1877 de grootste uitvinder aller tijden, Thomas Edison (hij schopte het tot het recordaantal van 1093 patenten), hoe hij zijn pas gerealiseerde fonograaf in een sprekende, 'intelligente' en vrouwelijke pop zou kunnen omtoveren. Hij zag eindelijk de mogelijkheid om het zwakke vlees door willige machines te vervangen! Edison ging als een bezetene aan de slag: de romp en ledematen bouwde hij van zes stukken tin en in het hart van zijn creatie kwam een state-of-the-art miniatuur-fonograaf te zitten. Nog geen jaar later liet hij de bezoekers van de wereldtentoonstelling in Parijs versteld staan van de eerste levensecht sprekende pop. Een unicum! De laatste barrière tussen mens en dier - sterker nog - tussen mens en ding, namelijk het vermogen te kunnen spreken, was eindelijk overwonnen. De realisatie van mechanische, levende wezens zou nu nog een kwestie van tijd zijn. Nog verder terug in de tijd, in de dertiende eeuw, spendeerde de monnik Albertus Magnus dertig jaar van zijn leven om een sprekende androïde te bouwen. De legende wil, zo schrijft Wood, dat zijn beroemdste leerling, Thomas van Aquino, zo schrok van de telkens 'Salve' groetende robot dat hij hem meteen met een ijzeren staaf in elkaar sloeg.

Wood rijgt zo een aantal beroemde en minder bekende voorbeelden van historische androïden en robots vlot achterelkaar, en de gedachte die dat onvermijdelijk oplevert is dat het moderne klaroengeschal over intelligente robots over enige tijd ook wel eens een plaatsje in het historische rariteitenkabinet zou kunnen krijgen.

Toen Wood deze historisch relativiserende gedachte uiteindelijk eens voorzichtig aan Brooks voorlegde, antwoordde hij haar met een zin die misschien de tijdloze wereld van de droom van machinale intelligentie het meest treffend illustreert: 'We live in this weird world - there is no past, everything is the present.'

Al willen ingenieurs als Brooks er niets van weten: de geschiedenis van het debat over de verhouding tussen machinale en menselijke intelligentie laat zien dat de onophoudelijke en in een razend tempo voortgaande stroom van technische uitvindingen al eeuwenlang nauwelijks nieuwe filosofische gezichtspunten heeft opgeleverd. Integendeel, het lijkt erop dat het raadsel van ons bewustzijn steeds dezelfde soort intuïties oproept. Daarbij neigen mensen ertoe het eigen vermogen tot redeneren te objectiveren in de machines die ze zelf maken. In plaats van voorspellingen te doen over hoe intelligent machines binnenkort wel of niet zullen zijn, is het daarom wellicht interessanter te kijken naar de

mechanismen waarop het verstand zichzelf al eeuwenlang telkens weer met nieuwe technische uitvindingen probeert te spiegelen, te imiteren en te verslaan.

**Sybe Rispens** schrijft vanuit Berlijn over wetenschap en techniek. Hij promoveert binnenkort op het proefschrift *Gears of Thought. Clocks, computers, and the rise of modern science*.

**Besproken boeken:**

**Rodney Brooks** en **Sandy Fritz** (eds.), *Understanding artificial intelligence. From the editors of Scientific American*. New York: Byron Preiss 2002, 160 blz., 10,14

**Rodney Brooks**, *Flesh and Machines. How Robots will change us*. New York: Pantheon Books 2002, 260 blz., 20,37 (In Nederlandse vertaling verschenen als *De kunstmatige mens. Hoe machines ons veranderen*. Amsterdam: De Bezige Bij 2002, 304 blz., 24,50

**Max Urchs**, *Maschine, Körper, Geist. Eine Einführung in die Kognitionswissenschaft*. Frankfurt am Main: Vittorio Klostermann 2002, 333 blz., 19,80

**Gaby Wood**, *Living Dolls. A magical history of the quest for mechanical life*. London: Faber and Faber 2002, 289 blz., 22,11