

# Wat computers nog steeds niet kunnen

Er is de laatste tijd ook in dit blad nogal wat te doen geweest over het brein en het zelfbewustzijn van mens en machine. Een van de weinigen die zich al vroeg duidelijk heeft uitgesproken over de onmogelijkheden van AI, is de Amerikaanse filosoof Hubert Dreyfus. In zijn boeken *What computers can't do*, *Mind over machine* en het meest recente *What computers still can't do*, heeft hij de AI-mythe doorgeprikt en aangegeven waar de grenzen van machines liggen.

Van de bekende maar controversiële Amerikaanse psycholoog B.F. Skinner is de uitspraak: "De kwestie is niet of machines kunnen denken maar of mensen dat doen." Ik heb iets tegen woordgebruik als 'een zelfdenkende telescoop', waar bedoeld is een telescoop die zichzelf automatisch kan bijstellen. Ik begrijp ook niet dat kranten die bol staan van de afbreekfouten door een computer, zelfs maar suggereren dat computers zouden kunnen denken. En als computers ook maar een greintje zelfbewustzijn zouden hebben, zouden ze zich rot moeten schamen.

## Hoger niveau

Maar goed, het opmerkelijke in de discussie over zelfbewustzijn bij mens en machine is dat zij naar mijn smaak ten onrechte naar een hoger niveau getild is. Zo'n jaar of twintig geleden was het de vraag of machines intelligent gedrag zouden kunnen nabootsen, dat wil zeggen kunstmatig intelligent zouden kunnen zijn. Het is inmiddels wel duidelijk dat dit praktisch onmogelijk is. Naast nogal wat nieuwe programmeertechnieken, is in mijn ogen een van de belangrijkste resultaten van alle AI-onderzoeken geweest dat men is gaan beseffen hoe onvoorstelbaar complex menselijke intelligentie in elkaar zit. Daarom vind ik het zo opmerkelijk dat men nu over zelfbewustzijn en computers is gaan praten.

De AI-wereld is van meet af aan gekenmerkt door twee soorten redeneringen. De eerste redenering berust op een belofte: omdat we deze eerste stap kunnen doen, zal een computer straks bepaald intelligent gedrag gaan vertonen. Helaas horen we daarover later nooit meer iets. We horen het alleen als er opnieuw een eerste stap is gezet, waardoor een computer nog intelligenter zal kunnen worden. De overgang van intelligentie naar zelfbewustzijn is weer zo'n nieuwe eerste stap. Ik vergelijk dit soort redeneringen met die van een kind dat zegt: ik kan nu de ene voet voor de andere zetten, daarom kan ik straks de 100 meter in 8 seconden lopen. Om daarna met de verzekering te komen: omdat ik tijdens het kruipen een schaakstuk heb gegrepen, zal ik straks tijdens het hardlopen ook nog de wereldkampioen schaken verslaan.

## Principiële stellingname

De tweede redenering berust ook op een misverstand: omdat een computer met symbolen kan manipuleren, kan hij ook begrijpen wat die symbolen in de menselijke wereld betekenen. Neem het voorbeeld van een willekeurige computer, waarin een complete beschrijving van zijn ontwerp en functioneren is opgeslagen. Kan deze computer nu zichzelf begrijpen? Of met deze kennis iets leuks doen, iets wat hem niet door mensen is opgedragen, in welke vorm dan ook? Het antwoord op die vraag is natuurlijk dat dit in principe niet

onmogelijk is, zolang we de daarvoor benodigde menselijke betekenissen en wenselijkheden niet ook in die computer zouden inbrengen. En daarin schuilt het probleem: er zijn oneindig veel verschillende menselijke betekenissen en wenselijkheden, die we ons voor het grootste gedeelte niet eens bewust zijn. Die menselijke bewustwording is noodzakelijk om betekenissen en wenselijkheden expliciet te maken, wat nodig is om ze over te brengen in een computer; dit maakt het voorbeeld in de praktijk volslagen onuitvoerbaar.

Mede hierdoor vertoont de zoektocht naar kunstmatige intelligentie veel overeenkomst met de zoektocht van alchemisten naar methoden om goud te maken. Een van de weinigen die zich al vroeg duidelijk heeft uitgesproken over de onmogelijkheden van de kunstmatige intelligentie, is de Amerikaanse filosoof Hubert Dreyfus. In zijn boeken *What computers can't do* (1972) en *Mind over machine* (1986) heeft hij de AI-mythe doorgeprikt en aangegeven waar de grenzen van machines liggen. In 1992 is een herziene versie van deze boeken verschenen, met een nieuwe uitgebreide introductie die zijn principiële stellingname goed verduidelijkt. De inleiding van het boek windt er geen doekjes om: "Deze editie betekent niet alleen een andere uitgever en een kleine verandering in de titel; zij betekent ook een verandering van status. Het boek biedt niet een controversiële stellingname in een lopende discussie, maar een beeld van een voorbije periode." Een pikant detail is dat de nieuwe uitgever, MIT Press, vroeger veel AI-boeken publiceerde. Het MIT (Massachusetts Institute of Technology) is in de AI-wereld een belangrijk onderzoekscentrum. Nog steeds vinden hier veel activiteiten plaats.

### **Kritische analyse**

De ideeën van Dreyfus zijn tot nu toe vrijwel geheel overstemd door de claims van AI-adepten. Het is te hopen dat zijn recentste publicatie hierin verandering brengt. Een uitgebreide introductie bij de MIT-editie en bij het herziene boek, zal zeker veel bijdragen aan de toegankelijkheid van zijn in aanleg nogal abstracte ideeën. Waarschijnlijk zal niet iedereen het eens zijn met zijn kritische analyse van de ontwikkelingsgeschiedenis van de AI-wereld, maar ik denk dat het vernieuwde boek veel stof tot nadenken biedt voor iedereen die geïnteresseerd is de grenzen van computers. De achterflap stelt: "Het boek *What Computers Can't Do* werd door vrijwel iedereen aangevallen maar in stilte bestudeerd. De redeneringen van Dreyfus zijn nog steeds provocatief en vragen onze aandacht nogmaals voor het unieke in de mens." In dit kader is het interessant dat de Amerikaanse overheid de jaren negentig heeft uitgeroepen tot *Decade of the Brain* en een groot onderzoeksprogramma is begonnen. Kortom, als u problemen heeft om een AI-project te verwezenlijken, als u wilt weten hoe het de kunstmatige intelligentie vergaan is, of als u gewoon geïnteresseerd bent in de vraag wat wij met computers nu wel en niet kunnen doen, dan biedt het boek van Dreyfus veel interessants.

Hein van Steenis, freelance medewerker van Computable (22 april 1994)

H.L. Dreyfus. *What computers still can't do*. The MIT Press (1992), ISBN 0262540673.