

De Ivoren Toren van de AI

Over-optimisme en zelf-reflectie binnen de AI

Alex de Landgraaf, 2003

Inhoudsopgave

Inleiding.....	3
Wat streeft de AI na? En waarom?.....	4
Waarom het eeuwige optimisme van de AI?.....	6
Wanneer wordt het middel heilig?	7
Zelfreflectie, of het gebrek daaraan?	8
Slotwoord.....	9
Bronnen.....	10

Inleiding

Het vakgebied AI en optimisme zijn nauwelijks van elkaar te scheiden. AI sinds het begin van de AI zijn geen bergen te hoog of ze zullen beklommen worden.

Echter, in elk gebied is het verstandig om af en toe afstand te nemen van het dagelijks doen en laten om eens objectief te reflecteren waar je mee bezig bent. In deze paper wil ik niet proberen om een geschiedenis te schetsen van de AI, of over alle technieken en voordelen die uit de AI zijn ontstaan uit te wijden. Ook wil ik niet de nieuwste mogelijkheden en plannen schetsen. In deze paper wil ik de AI een blik op zichzelf laten werpen, om mensen in dit vakgebied aan te sporen om de wereld zonder oogkleppen waar te nemen, maar bovenal om een gezonde discussie kritiek te ontwikkelen tegenover het vakgebied van het eeuwige optimisme.

Wat streeft de AI na? En waarom?

Het is januari 1956. Herbert Simon gaf les aan Pittsburgh's Carnegie Tech. Tijdens het begin van zijn les wiskundig modeleren verklaarde hij, dat hij, samen met 2 collega's, een machine had gebouwd die kon denken. Zijn programma kon logische theoriën ontdekken, iets dat tot dan alleen door mensen gedaan zou kunnen worden. Zijn belangrijkste bewijs was echter: het bewijs dat de menselijke brein helemaal niet zo bijzonder is...

Artificial Intelligence

n : *the branch of computer science that deal with writing computer programs that can solve problems creatively*; "workers in AI hope to imitate or duplicate intelligence in computers and robots" [syn: AI]

in-tel-li-gence

n.

1. *The capacity to acquire and apply knowledge.*
2. *The faculty of thought and reason.*
3. *Superior powers of mind.*

Dit lijkt 'eenvoudig' genoeg. AI heeft zichzelf tot doel gesteld om intelligentie te **imiteren** of **na te maken**. Dit geeft meteen de twee stromingen aan binnen de AI: de Weak AI, oftewel de imitatie van intelligentie, en de Strong AI, oftewel het namaken van intelligentie. Echter, een derde, meer menselijke doel is het begrijpen van de mens zelf, dit is het gebruiken van AI om door te krijgen wat de intelligentie van de mens inhoud.

Maar waarom de eerste twee doelen?

Waar komt die drang vandaan om een op zichzelf werkende machine te maken die intelligent lijkt? Is dit een vorm van hoogmoed, gevoed door de hedendaagse computertechnologie, om iets te creëren dat intelligentie bevat?

Nee, de drang om het niet-levende levend te maken bestaat naar mijn mening al sinds de oudheid. In die tijd zien we de eerste autonome systemen die, hoe ironisch het ook mogen zijn, dienst deden als magie van de goden. Daniel Crevier noemt hierbij een goddelijke beeld dat dienst deed bij een ceremonie om de volgende pharaoh in het oude Egypte te kiezen. Deze stak zijn arm uit en sprak wanneer de volgende pharaoh langs liep bij een ceremonie:

"Although it is likely that those who took part in the process, would-be pharaohs and onlookers alike, knew that a bit of theater was being put on, the procedure was taken seriously: in the Egyptian mind, the priest-cum-statue system added up to more than the sum of its parts, and embodied the god"

Het autonome spreekt tot de verbeelding, mensen zien hierin het goddelijke, het onbegrijpbare. In de Middeleeuwen waren het talloze radertjes en veertjes in de nieuwerwetse klokken, of trucages waarbij het lijkt alsof een pop zou kunnen schaken terwijl er eigenlijk een dwerg onder het bord al het denkwerk doet. Het maakt niet uit wat, maar zaken die niet verklaard kunnen worden spreken tot de verbeelding: kan een machine denken?

Naast deze fascinatie hebben we echter ook een tegenovergestelde reactie: de angst, dat deze creaties beter zullen zijn dan de mens. We hoeven alleen aan Frankenstein te denken, de creatie van aan elkaar genaaide lappen vlees, die door middel van bliksem tot leven komt en zijn maker doodt.

Soortgelijke verhalen en legenden doen al eeuwen de ronde, zoals de Golem die een Rabbi vermoedelijk zou hebben gemaakt door middel van mystiek en religieuze bezweringen. Echter, grote ongelukken gebeuren wanneer anderen de Golem proberen te controleren. Boven-natuurlijke krachten zijn nodig om zulke wezens te creëren, maar het gebruik van zulke krachten is niet voor de gewone mens bestemd. Daar komen tenslotte alleen ongelukken van...

Mary Shelly, de schrijfster van Frankenstein over haar boek:

"Frightful must it be; for supremely frightful would be the effect of any human endeavor to mock the stupendous mechanism of the Creator of the world"

Uiteindelijk zijn wij toch ook 'mechanismen', gemaakt door een hogere Maker?

Wij kunnen hieruit afleiden dat de wil om iets natuurlijk als intelligente, zichzelf voortbewegende machines te maken al zeer vroeg iets menselijks was. De Strong AI heeft zich tot doel gesteld om met de 'stupendous mechanism of the Creator of the world' te spotten.

Waarom het eeuwige optimisme van de AI?

We hebben nu gezien dat de AI niet zo onnatuurlijk is als dat het lijkt, de wil om iets levends te creëren is blijkbaar niet zo nieuw. Maar waarom lijkt de AI blind op zijn doel af te gaan? Waar komt dit blinde vertrouwen op de techniek vandaan, of kijkt de AI wel kritisch naar waar ze mee bezig zijn?

op-ti-mism

n.

1. *A tendency to expect the best possible outcome or dwell on the most hopeful aspects of a situation:* "There is a touch of optimism in every worry about one's own moral cleanliness" (Victoria Ocampo).

In een interview uit 1994 met Herbet A. Simon, een van de grondleggers van AI in de jaren '60, beweert hij dat het "over-optimisme binnen de AI" een mythe is die begonnen was door Dreyfus in de jaren 70. "They've made great progress to the top of the tree, but they're not at the moon yet" is een uitspraak van Richard Bellman, ook iemand die felle kritiek op de AI had.

Simon weerlegde hun stellingen door erop te wijzen dat AI nog jong is (het bestaat pas zo'n 40 jaar in 'officiële' vorm) en dat in ieder vakgebied er nu eenmaal periodes zijn waarin er weinig vooruitgang wordt geboekt.

Verder in de interview wordt gevraagd waarom de meeste voorspellingen van onderzoekers zo roos-kleurig zijn. Simon antwoordt dat zijn voorspellingen uit zijn gekomen die hij maakte in 1957 wat betreft AI. Het enige is, dat hij verwachtte dat een schaakcomputer binnen 10 jaar wereldkampioen zou zijn. Verderop geeft hij toe dat het wel 40 jaar zal duren. Hij zat er wat die voorspelling betreft flink naast, maar hij stelt dat 3 andere voorspellingen wel zijn uitgekomen. Het probleem is alleen dat deze voorspellingen zo vaag zijn dat ze makkelijk uit waren gekomen, alleen het schaken was glashelder en kwam niet uit. Duidelijk een onvervalste optimist; hij beweert dat hij geen beslissing zou kunnen bedenken die hij niet in handen van een (geavanceerde) AI systeem zou kunnen overlaten. Vroeger hoopte hij, toen hij in slecht weer landde op een vliegveld dat er een bekwame piloot aan het stuur was; nu hoopt hij, dat liever een bekwame computer het landen doet. Hij vindt computers betrouwbaarder dan mensen.

Dus, aan de ene kant beweert deze AI'er dat er geen over-optimisme is, dat het een verzinsel van Dreyfus is. Aan de andere kant vertrouwt hij meer op machines dan de mens. Het vroegere over-optimisme wordt goed gepraat en de trein genaamd AI dendert verder. Wat is het eindstation?

Wanneer wordt het middel heilig?

De AI als een technologische religie?

"The mechanical arts are man's links with the Divine, their cultivation a means of salvation". De Filosoof John Scotus Erigena uit de negende eeuw schetst hier de technologie en wetenschap uit die tijd, ze zijn een manier om dichter bij het Goddelijk te komen.

Religion

n 1: *a strong belief in a supernatural power or powers that control human destiny*; "he lost his faith but not his morality"

2: *institution to express belief in a divine power*; "he was raised in the Baptist religion"; "a member of his own faith contradicted him"

In "The Religion of Technology" gaat David Nobel in op de technologie van vroeger en het heden, waarbij hij elke keer links legt naar de religie en laat zien, dat beiden diep met elkaar verbonden zijn. Ook de AI wordt hierin belicht. AI is een vakgebied, waarin de mens droomt om iets superieurs dan de mens zelf te maken, door iets levenloos het goddelijke te schenken: bewustzijn. De droom om iets levens uit dood materiaal te creëren zou diepgeworteld zijn in de Middeleeuwse alchemie, iets dat gedeeltelijk wel klopt, maar hierboven heb ik laten zien dat deze drang veel verder gaat. Eeuwige leven zou binnen handbereik zijn, het Paradijs is dichtbij.

Is AI niet meer dan een droom, niet meer dan een religie waar men zich aan vastklampt? Is het niet meer dan een vorm van ultieme grootheidswaan om te denken dat iets wonderbaarlijks als de mens nagemaakt kan worden, dat bewustzijn iets is dat door middel van regels en symbolen gevangen kan worden? Is de mens werkelijk niet meer dan een machine, een speciale machine welteverstaan, waarvan het verstand gevangen kan worden binnen een computer?

In dat licht is AI meer op te vatten als een secte, een die tot de jaren zeventig een gigantische bloeiperiode zag, maar die gevangen werd door de onmogelijke opdracht die het zich opgedragen had. Naarmate dat steeds duidelijker werd, gingen de onderzoekers zich richten op systemen die een deel-probleem moesten oplossen, expert-systemen werden gebouwd om real-world problemen op te lossen.

Zelfreflectie, of het gebrek daaraan?

Waar zijn de Critics?

Toen tegen de jaren '70 onderzoekers aan de expert-systemen begonnen te bouwen, hoopten ze dat ze deze konden uitvergroten naar de intelligentie van de mens. Het zou alleen een kwestie van tijd worden voordat alle informatie van de gewone mens op te slaan zou zijn. Zelfs nu nog zijn er onderzoekers die tegen het 'scalen' aankijken als het belangrijkste probleem voor de AI.

Dendert de AI trein dan gewoon door? Waar is de kritiek tegen de AI? Dit komt vooral vanuit de filosofische hoek, voornamelijk omdat het onmogelijk zou zijn om echte intelligentie te evenaren. Dreyfus, Taube, Searle, ze kwamen allemaal uit deze hoek om de AI'ers op de grond te houden, al waren sommige filosofen best geïnteresseerd in hoe de AI'ers te werk gingen. Dreyfus: "I know from experience that challenging these assumptions will produce reactions similar to those of an insecure believer when his faith is challenged". De andere kant kent talloze namen uit de AI, met als stamvader van de AI de al eerder besproken Herbert Simon. Hij verweert zich met: "You don't get very far arguing with a man when his faith is challenged, and these are essentially religious issues to the Dreyfuses and Weizenbaums of the world". Beide kampen verklaren dat de ander religieuze vooroordelen heeft. Is het niet zo dat beide kampen een andere religie aanhangen, een ander wereldbeeld, die niet met elkaar samen te voegen is? Beiden zien intelligentie en bewustzijn als iets volkomen anders, maar is het niet naïef om te denken, dat de ander van zijn geloof zal vallen?

Nu is er een probleem met deze twee-kampen-strijd: de AI zal zich afzonderen, kritiek zal niet worden geuit op de weg die gekozen is. Juist vanwege deze strijd worden de believers en de non-believers tegen elkaar opgesteld.

Echter, er zijn uitzonderingen. Joseph Weizenbaum is een belangrijke onderzoeker in de AI (maker van onder andere ELIZA, een 'psychiater expert'. Hij twijfelde openlijk tegen de stelling dat AI mogelijk is en vond dat het op zijn minst moreel twijfelachtig zou zijn om deze te maken als het mogelijk zou zijn. Het gebruik van AI zou gigantische consequenties met zich meebrengen, maar hij vergelijkt het gebruik van techniek voor on-morele doeleinden als iets wat de onderzoeker rationaliseert. De wetenschap wordt volgens Weizenbaum gezien als iets autonooms, iets dat uit zichzelf vooruit gaat. "It leads to the position that 'If I don't do it, someone else will.'" Hij trekt dit door naar de normale ethiek: "People will be murdered; If I don't do it, someone else will". Dit is natuurlijk een fatalistische instelling die geen (echte) onderzoeker zou kunnen innemen. "It depends how one uses it. [...], we scientists cannot know how it is going to be used. So therefore we have no responsibility". Echter, we leven in een echte maatschappij. Je kan niet stellen dat, als je werkt aan atoom energie, raketten of lasers, dat deze uitvindingen niet voor slechte doeleinden zullen worden gebruikt. Uiteindelijk kan elke belangrijke ontdekking van de laatste eeuw worden gebruikt voor slechte doeleinden. Weizenbaum vreest dat de nieuwste generatie wetenschappers dit niet doorhebben, maar hij hoopt dat hij ongelijk heeft.

Slotwoord

Zijn er mensen die proberen de ogen te openen van de AI'ers? Ja, ze zijn er. Maar mensen zoals Weizenbaum lijken me dun gezaaid binnen het vakgebied AI (en Informatica). Kritiek op en binnenuit het vakgebied lijkt echter weggedrukt te worden. Weggewuifd door mensen zoals Simon die het afdoet als religieuze geratel.

Wanneer opent de AI haar ogen en durft ze de consequenties te aanschouwen van waar zij mee bezig is? Wanneer wordt de AI wakker? Onderzoek is pas wetenschap, wanneer deze objectief wordt bedreven. Het gebruiken van de noodrem in de AI-trein die alleen maar doordendert, zal niet mogelijk zijn. Maar misschien wordt het tijd om wat vaart te minderen, zodat we de afgrond tenminste zien aankomen?

Bronnen

Roger C. Schank, Where's the AI?

Dunn, J. S, Thinking Machines

AI and Ethics,
<http://www.aaai.org/AITopics/html/ethics.html>

Nobel, D. F. The Religion of Technology: The Divinity of Man and the Spirit of Invention.

Noble, D. F. 1977. America by Design: Science, Technology and the Rise of Corporate Capitalism.

The Internet Encyclopedia of Philosophy
<http://www.utm.edu/research/iep/a/artintel.htm>

An Interview with Weizenbaum
<http://www-tech.mit.edu/V105/N16/weisen.16n.html>

Herbert A. Simon: Thinking Machines
<http://www.omnimag.com/archives/interviews/simon.html>

Essays on the Philosophy of Technology
<http://commhum.mccneb.edu/PHILOS/techessay.htm>

The American Heritage Dictionary of the English Language,
Fourth Edition

WordNet 1.6, 1997 Princeton University