

Inhoud

Proloog	9
---------	---

Deel I Zoektocht naar Kunstmatige Intelligentie

1	Calculus	17
2	Hersenen versus computers	26
3	De eerste slimme machines	37

Deel II Natuurlijke en Kunstmatige Intelligentie

4	Hoe wij intelligent werden	53
5	Doel van het leven	64
6	De zintuigen van een computer	74
7	Een stroom van informatie	82
8	Het kindprogramma	100
9	Machine learning	110
10	Wat willen we leren?	121
11	Het trainen van computers	138
12	Intelligentie die we zelf niet begrijpen	147
13	Alles is abstract	154
14	De motor van intelligentie	175

Deel III Toekomst van (Kunstmatige) Intelligentie

15	De drijvende krachten van AI	195
16	Kan een computer alles wat een mens kan?	217

17	Het ontwerp van AI	230
18	De schaal van intelligentie	242
19	De veiligheid van AI en van mensen	252
20	Het moeilijke probleem	269
	Epiloog	283
	Noten en referenties	287

Proloog

Cogito, ergo sum. Ik denk, dus ik ben. De beroemde uitspraak van de Franse filosoof René Descartes (1596-1650) legt de vinger op de zere plek. Denken is niet zomaar een eigenschap van de mens; denken definieert ons. Niet voor niets noemen we onze soort *homo sapiens*: de denkende mens. Ons vermogen tot denken, onze *intelligentie*, is onze meest karakteristieke eigenschap.

Intelligentie is een van de fascinerendste fenomenen in dit universum. Ik heb me in ieder geval altijd verwonderd waarom we kunnen wat we elke dag weer doen. Wat gebeurt er in ons brein als we discussiëren, schaken of plannen maken voor morgen? Hoe kunnen we op school leren over de meest uiteenlopende vakgebieden, van scheikunde tot economie en literatuur? En waarom kan ik deze tekens überhaupt opschrijven, en hebben ze voor jou nog betekenis ook? Het zijn vermogens die – voor zover we weten – nergens anders in het universum te bekennen zijn. Wat *is* dit voor proces, dat blijkbaar plaatsvindt in ons brein?

Intelligentie is ook een ongrijpbaar fenomeen. Het vindt pal onder onze ogen plaats, maar het blijkt wetenschappelijk lastig de vinger achter het onderliggende mechanisme te krijgen. Over de jaren zijn weliswaar vele wetenschappelijke puzzelstukjes gevonden, bijvoorbeeld via psychologisch en neurowetenschappelijk onderzoek naar de werking van ons brein. Maar intelligentie is zo'n complex proces, dat wetenschappers nog altijd bezig zijn de volledige puzzel te leggen.

Nog voordat we een volledig antwoord hadden, ontwikkelde de wetenschap zich ook in andere richtingen. Een van deze paden

leidde naar *kunstmatige intelligentie*, ook wel artificiële intelligentie, doorgaans afgekort tot ‘AI’. Ook dit is uiteraard een vorm van intelligentie, maar dan in een niet-levend medium. Deze technologie komt, zeker in de eenentwintigste eeuw, steeds sneller onze maatschappij binnen. Onze digitale assistenten worden slimmer en gaan een almaar grotere rol spelen in ons dagelijks leven, terwijl AI op de achtergrond revoluties ontketent in vrijwel alle wetenschappen – van nieuwe medicijnen tot betere computerchips en geavanceerde bedrijfsvoering. AI kruipt in de poriën van onze samenleving.

Met de nieuwe mogelijkheden van AI volgen ook veel nieuwe vragen. Gevorderde intelligentie leek immers een uniek menselijke eigenschap, maar nu blijken computers er ook steeds meer toe in staat. Maar hoe kan een computer eigenlijk slim zijn? *Wat gebeurt daar vanbinnen?* Dat is niet alleen een interessante vraag, maar ook een urgente. Iedereen in onze maatschappij komt in aanraking met AI. Maar om de technologie veilig te gebruiken moeten we wel snappen hoe deze echt werkt – en wat haar kwetsbaarheden zijn.

De technologische veranderingen volgen elkaar ook in hoog tempo op. Dat roept bij veel mensen terecht vragen op over de toekomst: wat staat ons te wachten? We moeten bedenken waar het pad van AI naartoe zou kunnen gaan, en van welke factoren dat proces afhankelijk is. En uiteraard is de vraag ook waar wij zelf staan in dat proces: hoe willen (of moeten) we onze gezamenlijke toekomst met AI vormgeven?

En dan brandt nog die andere vraag. Onze eigen intelligentie is immers fascinerend, maar ook nog altijd mysterieus. Als computers steeds beter in staat blijken delen van onze intelligentie na te bootsen, wat zegt dat dan over ons eigen denken? In hoeverre is de zoektocht naar een slimme computer ook een zoektocht naar onszelf? AI roept aan alle kanten cruciale vragen op – over haar werking, toekomst, en relatie met onszelf – die we in dit boek gaan onderzoeken.

Onderzoek naar intelligentie

Onze wonderbaarlijke intelligentie komt voort uit een even wonderbaarlijk orgaan: onze hersenen. Zo versturen de zenuwcellen in je brein op dit moment signalen met snelheden tot 400 kilometer per uur – als de snelste sportauto's. En ze versturen niet één zo'n signaal – elke seconde vliegen er miljarden signalen door je hoofd. Dit hele proces vindt plaats in een volume van nog geen anderhalve liter – in de orde van een pak melk. Maar dit brein maakt zoveel mogelijk. Het stuurt je spieren en hartslag als je gaat sporten, laat je communiceren met anderen, kan verleden onthouden én nadenken over de toekomst. Een van de meest complexe bouwwerken van dit universum zit verstopt in je hoofd.

Uiteraard gingen wetenschappers onderzoek doen naar dit orgaan. Twee vakgebieden hielden zich hier hoofdzakelijk mee bezig: de psychologie en de neurowetenschap. De psychologie bestudeert intelligentie en het menselijk denken doorgaans vanuit een *top-down*-benadering: het breekt het geheel (intelligentie) op in een reeks subfuncties en probeert te beschrijven hoe deze werken en met elkaar interacteren. Psychologen bestuderen bijvoorbeeld ons vermogen tot plannen, leren, aandacht en redeneren, en de manier waarop deze op elkaar ingrijpen.

In de neurowetenschap wordt meer naar de fysieke kant van intelligentie gekeken. Hierbij wordt onder andere de functie van verschillende gebieden in de hersenen in kaart gebracht. Zo ontdekten we de taalcentra, gelegen aan de zijkant van je brein, die verantwoordelijk zijn voor je begrip en productie van taal. Op kleinere schaal wordt veel onderzoek gedaan naar *neuronen* (zenuwcellen), de belangrijkste bouwstenen van onze hersenen. Daarbij is bijvoorbeeld veel ontdekt over de manier waarop neuronen signalen voortgeleiden en hoe ze zich aanpassen door onderlinge interactie. Dit proces is meer *bottom-up*, waarbij we eerst de kleine bouwsteentjes in kaart brengen, om van daaruit het volledige proces te doorgronden.

Midden twintigste eeuw ontstaat een derde onderzoeksrichting naar intelligentie: kunstmatige intelligentie. Na de ontwikkeling van de computer zien onderzoekers dat dit apparaat geschikt zou kunnen zijn om intelligentie na te bouwen. Maar hoe pakken we dat aan? Hoe kunnen we het – ogenschijnlijk ondoorgroendelijke – proces dat zich in ons brein afspeelt naar een computer verplaatsen?

Onderzoekers uit allerlei vooropleidingen, zoals de informatica, wiskunde, filosofie, natuurkunde en statistiek, gaan zich op dit vraagstuk richten. De uitdaging is natuurlijk ook prachtig. Mensen zijn immers slim – een fenomeen dat we doorgaans uitdrukken via het ‘IQ’ (intelligentiequotiënt). Maar zijn we ook zo slim dat we onze eigen slimheid kunnen nabouwen? Kunnen we *van IQ naar AI*? Kunstmatige intelligentie voelt als de ultieme proeve van slimheid – het potentiële kroonstuk van onze eigen intelligentie.

Over de jaren worden zo allerlei theorieën over de werking van intelligentie geïmplementeerd op computers. Daarbij komt ook een belangrijk voordeel van kunstmatige intelligentie naar voren. In computers kunnen we een theorie over intelligentie namelijk *testen*. We kunnen de methode (proberen te) implementeren en kijken of deze daadwerkelijk werkt: geeft de computer slimme antwoorden, lukt het een robot een taak voor ons op te lossen?

Het belang van dit experimenteel kunnen testen van een theorie wordt in mijn ogen vaak onderschat. De Engelsen hebben hiervoor een goed spreekwoord: *the proof of the pudding is in the eating*. Oftewel: de praktijk wijst uit of iets daadwerkelijk goed werkt. Het proeven van de pudding is denk ik ook van toepassing op intelligentie: pas als je in staat bent om intelligentie na te bouwen heb je het echt begrepen, en vice versa. Of, zoals natuurkundige en Nobelprijswinnaar Richard Feynman ooit zei: ‘What I cannot create, I do not understand.’

Proces

Door het experimenteel testen van theorieën ontwikkelde de AI-gemeenschap ook een nieuw perspectief op intelligentie. Om een computer iets intelligents te laten doen heeft deze namelijk gedetailleerde instructies nodig. Computers vragen om een exact stappenplan – een algoritme. Hierdoor gingen AI-onderzoekers intelligentie automatisch als een *proces* beschouwen. Ze gingen zoeken naar de bouwstenen (de instructies in het stappenplan) die tezamen het complexe bouwwerk (van intelligentie) mogelijk maken.

Over de jaren werd zo duidelijk welke bouwstenen in computers daadwerkelijk in staat zijn aspecten van menselijke intelligentie na te bootsen – en zeker ook welke niet. Die succesvolle bouwstenen zijn op zichzelf al fascinerend, maar roepen ook vragen op over onze eigen intelligentie. Natuurlijk is er geen garantie dat elk succesvol mechanisme uit een computer ook een rol speelt in ons brein – en ook in computers zijn we er nog lang niet uit. Maar als een AI-methode geïnspireerd is op ons brein en vervolgens significante aspecten van menselijke intelligentie begint na te bootsen, lijkt er alle reden beide werelden naast elkaar te leggen. Dat zullen we dan ook doen in dit boek: we gaan niet alleen van IQ naar AI – maar ook weer terug.

Daarmee zijn we klaar om te gaan denken over het denken – van computers en van onszelf. We gaan het doen, maar het is natuurlijk een opmerkelijk proces. Een tekst als deze is immers een soort cirkel: je verwerkt de inhoud met het proces waar je juist over leest. Na deze zin vuren de neuronen in je hersenen bijvoorbeeld gegarandeerd over het feit dat ze zelf aan het vuren zijn. Het vermogen om te denken over je eigen denken, om via denken zelfs de processen van je eigen denken te ontrafelen, is misschien wel het wonderbaarlijkst van allemaal. Laten we – met die reflectie in het achterhoofd – desondanks een poging wagen.

Deel I

Zoektocht naar Kunstmatige Intelligentie

Calculamus

Als AI-onderzoeker ervaar ik vaak dat mensen een grote fascinatie hebben voor kunstmatige intelligentie. Wanneer ik op een verjaardag vertel wat voor werk ik doe, krijg ik vrijwel zonder uitzondering geïnteresseerde vervolgvragen. Kun je uitleggen hoe de technologie echt werkt? Waar denk je dat AI in de toekomst toe gaat leiden? Zou een AI ooit bewust kunnen worden? AI *triggert* iets in mensen, en dat geldt waarschijnlijk ook voor jou – anders had je dit boek nu niet in handen.

De interesse van mensen in kunstmatige intelligentie is geen recente ontwikkeling: de mensheid denkt al veel langer na over de mogelijkheid van AI. Die geschiedenis loopt zelfs terug tot ver voor de ontwikkeling van de computer. Zo vinden we in de Chinese oudheid al de *Liezi*-tekst, waarin een handwerksman genaamd Yan Shi zich rond 1000 v.Chr. bij de koning meldt met een kunstmatige verschijning die we nu een robot zouden noemen:

‘De koning staarde verbaasd naar de gestalte. Het liep met snelle passen en bewoog zijn hoofd op en neer, op zo’n manier dat iedereen het voor een mens zou aanzien. De handwerksman raakte zijn kin aan, en hij begon te zingen, volledig zuiver. [...] Toen de voorstelling ten einde liep, knipoogde de machine en maakte avances naar de aanwezige dames, waarop de koning woedend werd en Yan Shi ter plekke zou hebben geëxecuteerd als deze niet, in doodsangst, de machine onmiddellijk aan stukken geslagen

had om hem te laten zien wat het werkelijk was. En inderdaad, het bleek slechts een constructie te zijn van leer, hout, lijm en lak, in verschillende kleuren wit, zwart, rood en blauw.'

Over de feitelijke juistheid van dit verhaal kun je uiteraard twijfelen. De tekst toont echter vooral dat de mens duizenden jaren geleden al bezig was met de mogelijkheid van kunstmatige intelligentie. De machine van Yan Shi is immers duidelijk mechanisch, gemaakt van leer, hout en lijm, maar vertoont allerlei intelligent gedrag, zoals zingen en flirten. En de Chinese oudheid is hierin niet uniek: soortgelijke mythes vinden we terug in andere oude beschavingen. Zo beschrijft de Griekse mythologie al *Talos*, een gigantische bronzen robot gemaakt door de goden om het eiland Kreta te beschermen. Blijkbaar is de mens al millennia bezig met de mogelijkheid van kunstmatige intelligentie, en zijn we onderdeel van een lange traditie. Maar waarom willen we dit eigenlijk – wat *drijft* de mensheid in haar zoektocht naar kunstmatige intelligentie?

Innovatie

Voor de eerste verklaring van onze interesse in AI moeten we nog verder de geschiedenis in. Als er één proces de rode draad vormt door de geschiedenis van onze soort, dan is het wel de eindeloze keten van *innovaties*. Stenen werktuigen, de beheersing van vuur, koken, kleding, landbouw, het wiel, de ploeg, schrift, operaties, de klok, buskruit, boekdrukkunst, de beheersing van elektriciteit, de stoommachine, de telefoon, de lamp, antibiotica – de keten van ontdekkingen is eindeloos. Over het algemeen stellen deze innovaties ons in staat om situaties te bereiken die eerder niet mogelijk waren: steviger gebouwen, hogere voedselproductie, nieuwe vormen van transport, betere medische zorg. De voortbouwende keten van menselijke innovatie verhoogt onze welvaart.

Onze interesse in AI is dan ook goed te verklaren als een vol-