

VAN **ARISTOTELES** TOT **ALGORITME**

© 2022 Guido van der Knaap, Amsterdam

© 2022 Boom uitgevers Amsterdam

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever. Voor zover het maken van kopieën uit deze uitgave is toegestaan op grond van artikelen 16h t/m 16m Auteurswet 1912 j° Besluit van 27 november 2002, Stb 575, dient men de daarvoor wettelijk verschuldigde vergoeding te voldoen aan de Stichting Reprorecht te Hoofddorp (Postbus 3060, 2130 KB, www.reprorecht.nl) of contact op te nemen met de uitgever voor het treffen van een rechtstreekse regeling in de zin van artikel 16l, vijfde lid, Auteurswet 1912. Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bijvoorbeeld een (digitale) leeromgeving of een reader in het onderwijs (op grond van artikel 16, Auteurswet 1912) kan men zich wenden tot de Stichting Uitgeversorganisatie voor Onderwijslicenties (Postbus 3060, 2130 KB Hoofddorp, www.stichting-uvo.nl).

No part of this book may be reproduced in any way whatsoever without the written permission of the publisher.

Boekverzorging: Marry van Baar

Afbeelding omslag: Laura Postma/Studio Ruw

ISBN 978 90 2443 260 8 | NUR 730

www.boomfilosofie.nl

www.boomuitgeversamsterdam.nl

Inhoud

Inleiding 7

1. De droom van Leibniz 11

Logica

2. De droom van Anderson 41

Epistemologie

3. Oude steegjes en kaarsrechte straten 73

Taalfilosofie

4. Ziel, brein en bewustzijn 101

Cognitiefilosofie

5. Een gedachte te veel 129

Ethiek

6. Macht en machines 159

Politieke filosofie

Tot slot 193

Dankwoord 201

Noten 203

Literatuur 210

Inleiding

Gebruik het gezonde verstand niet als een paraplu. Als je een kamer binnengaat om te filosoferen, laat het dan niet buiten, maar neem het mee naar binnen. De Oostenrijkse filosoof Ludwig Wittgenstein (1889-1951) gebruikte dit motto graag, naar verluidt ook tijdens een college over de grondslagen van de wiskunde. Bij deze colleges, gegeven in het voorjaar van 1939 in Cambridge, is de Brit Alan Turing (1912-1954) een van de kritische toehoorders.

Elf jaar na de colleges van Wittgenstein markeert diezelfde Turing met de vraag ‘Kunnen machines denken?’ het officiële begin van kunstmatige of artificiële intelligentie, kortweg AI. De vraag leidt hem naar een vraag-en-antwoordspel tussen een ondervrager, een mens en een machine. De deelnemers bevinden zich alle drie in een aparte ruimte, en de ondervrager moet erachter komen welke antwoorden van de machine zijn. De machine probeert zich voor te doen als een mens om zo de ondervrager om de tuin te leiden. Als dat lukt, is de machine volgens Turing intelligent.

Het vraag-en-antwoordspel gaat de geschiedenis in als de turingtest. De naamgever wilde er een helder criterium voor intelligentie mee formuleren. Dat blijkt echter lastig. Turing bespreekt zelf al verschillende argumenten tegen het idee dat een machine kan denken, ook als het lukt om de ondervrager te misleiden. Hij noemt onder meer het ontbreken van bewustzijn bij machines – een kwestie met een hoog filosofisch gehal-

te. Ruim zeventig jaar later is AI dergelijke theoretische discussies voorbijgesneld. De opmars van AI is veelal voortgestuwd door technici, programmeurs en grote bedrijven als Google. Niet door filosofen.

Kunstmatige intelligentie is meer dan een vraag-en-antwoordspel; er zijn toepassingen die uiteenlopen van vertaal-machines tot gepersonaliseerde zoekresultaten en gezichtsherkenning. Het roept de vraag op wat AI dan precies is. De Wetenschappelijke Raad voor het Regeringsbeleid (WRR) stelt bij de publicatie van zijn rapport *Opgave AI. De nieuwe systeem-technologie* dat AI 'allereerst moeilijk te definiëren' is.¹ Zelf kiest de WRR voor de definitie van een expertgroep bijengeroepen door de Europese Commissie. Volgens die definitie is AI 'het soort systemen dat intelligent gedrag vertoont door hun omgeving te analyseren en – met enige graad van autonomie – actie te ondernemen om specifieke doelen te bereiken'.² Dat is bewust een definitie met enige flexibiliteit; de WRR wenst ook ruimte te laten voor toekomstige toepassingen van AI. De tand des tijds bedreigt te starre definities.

Het is niet de bedoeling om hier te verzanden in definitiekwesties – dat is waar Turing bang voor is bij de vraag of machines kunnen denken, daarom introduceert hij een vraag-en-antwoordspel. Dit boek wil, nu AI steeds meer een plek in de samenleving krijgt, of op zijn minst nadrukkelijk op de deur klopt, een filosofisch licht laten schijnen op AI. Dat kan leiden tot een nieuw perspectief op toepassingen als Google Translate, gepersonaliseerde koopsuggesties en leren vanaf een *tabula rasa*, een onbeschreven blad.

Doordat AI zich laat gelden in de praktijk, beperkt een filosofische blik op AI zich niet tot de vraag of machines kunnen denken. De (mogelijke) inzet van gezichtsherkenning is in potentie een niet eerder vertoonde vorm van toezicht: nooit meer iemand onbespied. Spraakassistenten en vertaal-machines wagen zich aan de menselijke taal, terwijl ze ontegenzeggelijk van ons verschillen. Kunnen ze dan recht doen aan hoe de betekenis van taal tot stand komt? Kunstmatige intelligentie die leert

schaken, afbeeldingen kan classificeren en aanbevelingen doet op basis van data uit het verleden, beschikt over kennis, maar wat voor soort kennis is dat?

Deze vragen zijn geworteld in filosofische disciplines als politieke filosofie, taalfilosofie en epistemologie. Dit boek vertrekt vanuit deze disciplines – ook logica, ethiek en cognitiefilosofie komen aan bod – en laat nieuw licht schijnen op de fundamentele vragen die AI oproept. Daarbij komen uiteenlopende denkers aan bod, maar het biedt geen volledig overzicht. Net als het denken heeft ook dit werk zijn grenzen. Het wil vooral komen tot een beter (filosofisch) begrip van AI en de (on)mogelijkheden van kunstmatige intelligentie. Te beginnen in hoofdstuk 1 met de ontwikkeling van de logica, de filosofische basis van AI. In hoofdstuk 1 gaat het van Aristoteles tot de turingmachine, en van de ontwikkeling van AI in de tweede helft van de twintigste eeuw tot het spectaculaire succes van computers bij het populaire bordspel go.

‘Wat we niet denken kunnen,
dat kunnen we niet denken;
we kunnen dus ook niet *zeggen*,
wat we niet kunnen denken.’

Ludwig Wittgenstein,
Tractatus, §5.61

I

De droom van Leibniz

Logica

Alle Kretenzers zijn leugenaars, beweerde Epimenides rond 600 v.Chr. Wat hadden de bewoners van het eiland dat wij tegenwoordig vooral associëren met koning Minos, snikhete zomers en jongerenvakanties misdaan? Epimenides nam het de Kretenzers kwalijk dat zij Zeus als sterfelijk beschouwden, terwijl Zeus als oppergod natuurlijk onsterfelijk is. Dat lijkt slechts een religieuze twist in het oude Griekenland. De uitspraak van Epimenides krijgt echter een twijfelachtige status doordat Epimenides zelf ook Kretenzer is. Volgens de uitspraak is Epimenides dan zelf een leugenaar. Dat maakt de uitspraak dubieus, want die is gedaan door een leugenaar. En als de uitspraak niet klopt, dan zijn niet alle Kretenzers leugenaars, en is Epimenides mogelijk ook geen leugenaar. Met als gevolg dat de twijfel bij de uitspraak verdwijnt, waardoor alle Kretenzers ineens tóch leugenaars zijn.

Zulke paradoxale uitspraken geven aan hoe merkwaardig het denken in elkaar zit. In hetzelfde oude Griekenland is Aristoteles degene die het denken voor het eerst wil systematiseren met de ontwikkeling van logica. Natuurlijk is er de sokratische methode en dachten andere Griekse denkers na over waarheid, argumentatiestrategieën en paradoxen. Maar Aristoteles onderscheidt zich van zijn Griekse voorgangers doordat hij systematisch nadenkt over hoe je moet redeneren. De logica van Aristoteles is te vinden in zijn zes werken die bekendstaan als het *Organon*. Het betekent zoiets als ‘instrument’ of ‘ge-

reedschap', wat goed aangeeft hoe logica voor Aristoteles een manier van denken is die van pas komt bij het bedrijven van wetenschap.

Syllogismen zijn het tastbaarste en bekendste resultaat van de logica van Aristoteles. Een bekend voorbeeld is de volgende redenering: 'Alle Grieken zijn mensen, en alle mensen zijn sterfelijk, dus alle Grieken zijn sterfelijk.' Via syllogismen, en later andere methoden uit de logica, laat men zien dat de uitgangspunten van een redenering onvermijdelijk tot de bijbehorende conclusie leiden. Of, in een alternatieve formulering: het is onmogelijk dat de uitgangspunten waar zijn en de conclusie onwaar is.

Bij de redenering van zojuist zijn de uitgangspunten 'alle Grieken zijn mensen' en 'alle mensen zijn sterfelijk'. Daaruit volgt de conclusie dat alle Grieken sterfelijk zijn. Het is onmogelijk dat deze uitgangspunten, in jargon 'premissen', waar zijn en de conclusie onwaar is. Ga maar na: stel dat deze conclusie onwaar is, dan is er dus een Griek die niet sterfelijk is. Volgens de eerste premisse is die Griek echter een mens, en volgens de tweede premisse zijn alle mensen sterfelijk. Veronderstellen dat de conclusie onwaar is, is dus in tegenspraak met de premissen. Omdat we aannemen dat de premissen waar zijn, moet de conclusie ook waar zijn. Gegeven de premissen blijkt het onmogelijk dat een Griek niet sterfelijk is.

Deze redenering bestaat uit drie universele, bevestigende uitspraken: er wordt achtereenvolgens iets beweerd over alle Grieken, alle mensen en alle Grieken. Een redenering met deze vorm staat bekend onder de naam 'Barbara'. En bij logica is de vorm van een redenering cruciaal. Men kan probleemloos 'Grieken' door 'beren' en 'mensen' door 'dieren' vervangen; voor de logica verandert de aard van de redenering niet. Vergelijk het met elementaire wiskunde. Stel, Angela krijgt eerst twee taarten en daarna nog eens drie taarten. De wiskundige conclusie is dan dat zij $2 + 3 = 5$ taarten heeft. Deze conclusie komt tot stand door het toepassen van de wiskundige regels voor de getallen 2 en 3 en het plusteken, en niet door in

de echte wereld te gaan kijken naar hoeveel taarten Angela op dat moment heeft. Net als in de logica heeft een wiskundige conclusie dus niet per se iets te maken met de feitelijke situatie. De conclusie (het getal 5) is slechts het gevolg van de premissen $(2 + 3)$.

De logica van Aristoteles is daarom een eerste aanzet om te komen tot een wiskundig systeem waarin het onmogelijk is om van ware uitgangspunten tot een onware conclusie te komen. De potentie voor de ontwikkeling van kunstmatige intelligentie is enorm. Het idee is simpel: geef een computer alleen informatie die waar is, plus een goed werkend logisch systeem, en door die combinatie kan een computer zelf nieuwe conclusies trekken. Daarmee ligt het startpunt van de relatie tussen filosofie en kunstmatige intelligentie in de logica, het onderdeel van de filosofie om het denken te systematiseren.

Rekenen en denken

Het is echter een lange weg van de aristotelische logica naar de logica uit de tijd van Alan Turing. Niettemin betitelt Immanuel Kant (1724-1804) Aristoteles nog als begin- en eindpunt van de logica. Pas in de negentiende eeuw zal de logica nieuwe grote stappen voorwaarts maken. Dat weerhield denkers als René Descartes, Thomas Hobbes en Gottfried Leibniz er overigens niet van om in de eeuwen daarvoor al te speculeren over machines die kunnen denken. John Haugeland noemt Thomas Hobbes (1588-1679) in zijn *Artificial Intelligence. The Very Idea* zelfs 'de grootvader van AI'.¹ In het bijzonder vanwege Hobbes' stelling dat (rationeel) denken hetzelfde is als het maken van berekeningen. Zoals rekenen systematisch getallen bij elkaar optelt of van elkaar aftrekt, zo gaat het met denken ook, maar dan met speciale deeltjes in het brein. Evenals rekenen is denken gebaat bij het volgen van regels, aldus Hobbes.

Denken is dan het systematisch optellen en aftrekken van deeltjes in het brein. De sommen op papier en de bijbehorende