

MAX BENNETT

**EEN KLEINE
GESCHIEDENIS VAN
INTELLIGENTIE**

DE EVOLUTIE VAN
HET MENSELIJK BREIN
EN DE OPKOMST VAN AI

Vertaling Annemie de Vries

HarperCollins



Voor het papieren boek is papier gebruikt dat onafhankelijk is gecertificeerd door FSC® om verantwoord bosbeheer te waarborgen.
Kijk voor meer informatie op www.harpercollins.co.uk/green.

HarperCollins is een imprint van Uitgeverij HarperCollins Holland, Amsterdam.

Copyright © 2023 Max Solomon Bennett
Oorspronkelijke titel: *A Brief History of Intelligence*
Copyright Nederlandse vertaling: © 2024 HarperCollins Holland
Vertaling: Annemie de Vries
Omslagontwerp: Brian Moore
Bewerking: Pinta Grafische Producties
Omslagbeeld: © Getty Images
Foto auteur: © Gary O. Bennett
Zetwerk: MatZet B.V., Huizen
Druk: ScandBook UAB, Lithuania, met gebruik van 100% groene stroom

ISBN 978 94 027 1349 7
ISBN 978 94 027 6933 3 (e-book)
NUR 320
Eerste druk februari 2024

Het citaat van Darwin op blz. 7 is afkomstig uit: Hellemans, H. (2000). *Over het ontstaan van soorten*. Uitgeverij Nieuwezijds.

Originele uitgave verschenen bij William Collins, een imprint van HarperCollins-Publishers.

HarperCollins Holland is een divisie van Harlequin Enterprises ULC.
® en™ zijn handelsmerken die eigendom zijn van en gebruikt worden door de eigenaar van het handelsmerk en/of de licentienemer. Handelsmerken met ® zijn geregistreerd bij het United States Patent & Trademark Office en/of in andere landen.

www.harpercollins.nl

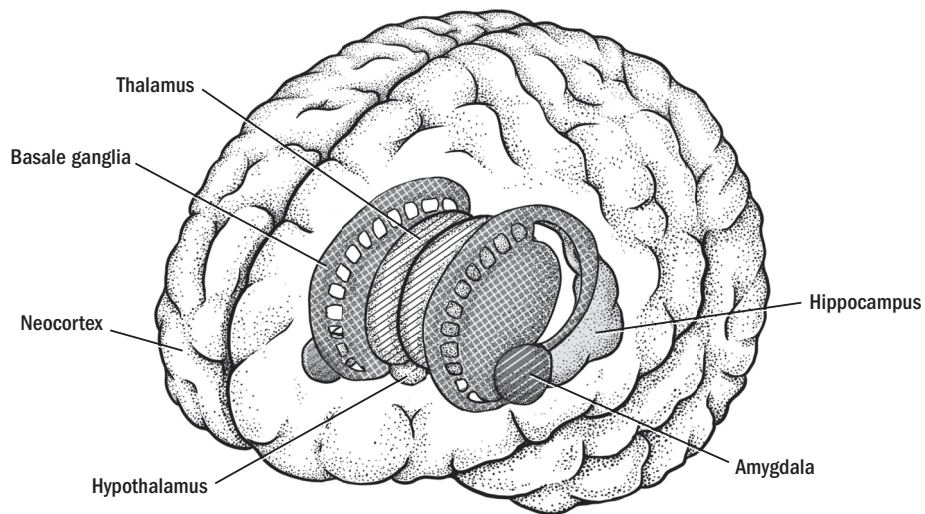
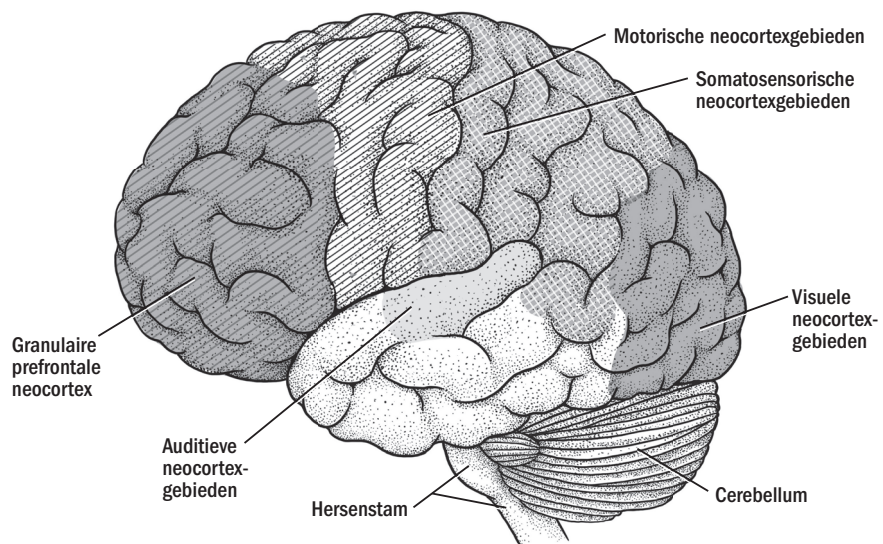
Niets uit deze uitgave mag openbaar worden gemaakt door middel van druk, fotokopie, internet of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van de uitgever. Het e-book is beveiligd met zichtbare en onzichtbare watermerken en mag niet worden gekopieerd en/of verspreid.

Inhoud

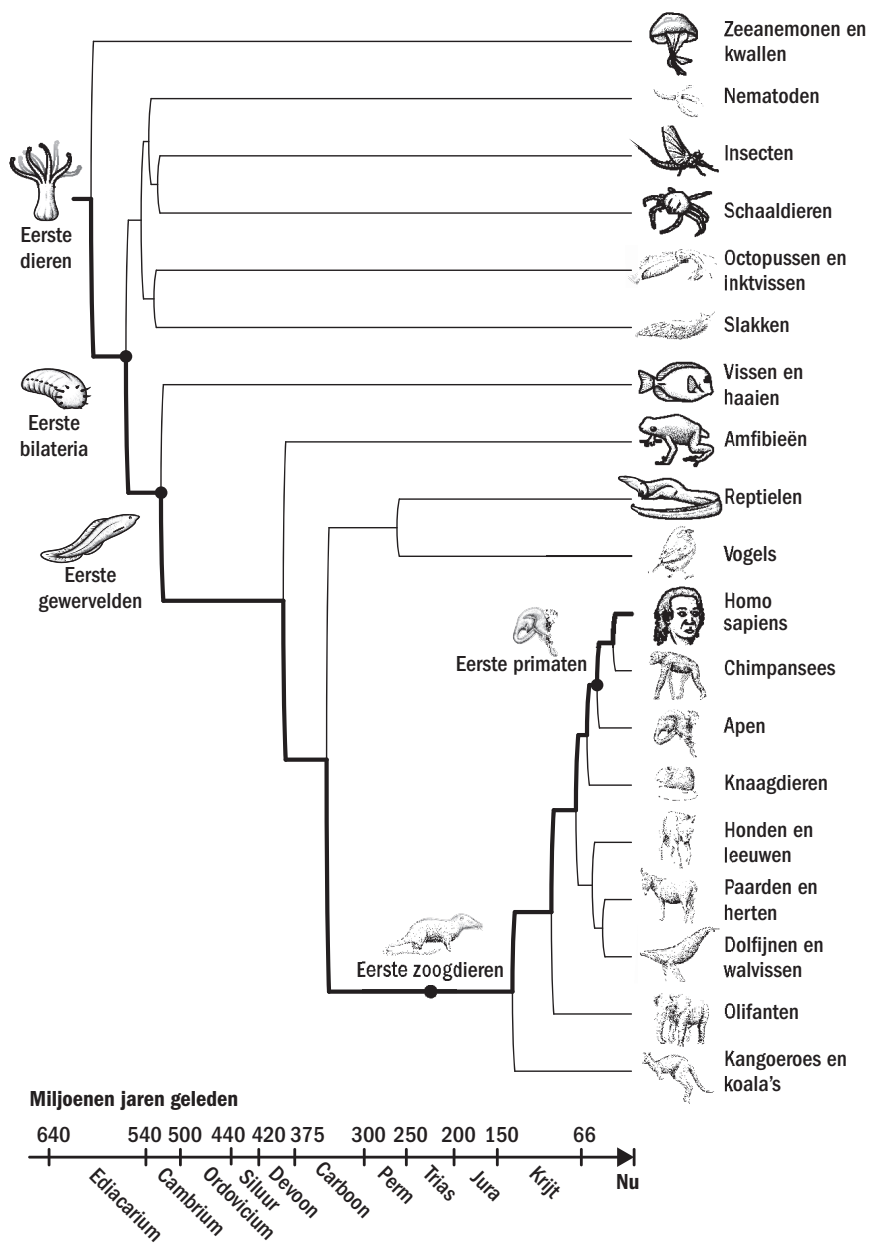
De basisanatomie van het menselijk brein	11
Onze evolutionaire afkomst	12
Inleiding	15
1: De wereld voordat er hersenen waren	31
Doorbraak #1: Sturen en de eerste bilateria	
2: De geboorte van goed en kwaad	59
3: De oorsprong van emotie	76
4: Associëren, voorspellen en het begin van leren	94
Doorbraak #2: Bekrachtigen en de eerste gewervelden	
5: De Cambrische explosie	113
6: De evolutie van <i>temporal difference learning</i>	123
7: De problemen van patroonherkenning	143
8: Waarom het leven nieuwsgierig werd	164
9: Het eerste model van de wereld	168
Doorbraak #3: Simuleren en de eerste zoogdieren	
10: De neurale duistere middeleeuwen	179
11: Generatieve modellen en het mysterie van de neocortex	190
12: Ratten in het rijk der verbeelding	212
13: <i>Model-based reinforcement learning</i>	227
14: Het geheim van afwasrobots	249

Doorbraak #4: Mentaliseren en de eerste primaten	
15: De wapenwedloop om politieke gewiektheid	265
16: Hoe je een model van andermans geest maakt	283
17: Apenhamers en zelfrijdende auto's	298
18: Waarom ratten geen boodschappen kunnen doen	315
Doorbraak #5: Spreken en de eerste mensen	
19: Op zoek naar het unieke van de mens	329
20: Taal in het brein	345
21: De volmaakte storm	358
22: ChatGPT en het venster naar de geest	381
Conclusie: De zesde doorbraak	397
Dankwoord	404
Begrippenlijst	408
Bibliografie	411
Verantwoording afbeeldingen, foto's en figuren	413

De basisanatomie van het menselijk brein



Onze evolutionaire afkomst



Inleiding

In september 1962, midden in het wereldwijde rumoer van de ruimtewedloop, de Cubaanse rakettencrisis en het pas verbeterde poliovaccin, was er nog een mijlpaal in de geschiedenis van de mensheid die minder aandacht kreeg, maar misschien even belangrijk was: in dat najaar van 1962 voorspelde we de toekomst.

Op de sinds kort kleurige Amerikaanse tv-schermen verscheen de eerste aflevering van *The Jetsons*, een tekenfilm over een gezin dat honderd jaar verder in de toekomst leefde. Het programma was, onder het mom van een tv-serie, eigenlijk een voorspelling: hoe mensen in de toekomst zouden leven, welke technologieën ze in hun zak zouden meedragen en in hun huis zouden neerzetten.

The Jetsons voorspelde videobellen, flatscreen-tv's, mobiele telefoons, 3D-printen en smartwatches; allemaal vormen van technologie die in 1962 onvoorstelbaar waren maar nu alomtegenwoordig zijn. Maar één technische uitvinding uit *The Jetsons* hebben we nog niet, één futuristisch kunststukje is nog niet tot bloei gekomen: de autonome robot Rosey.

Rosey verzorgde het gezin Jetson, ze paste op de kinderen en deed het huishouden. Toen zoontje Elroy van zes het moeilijk had op school, hielp Rosey hem met zijn huiswerk. Toen de vijftienjarige dochter Judy hulp nodig had om haar rijbewijs te halen, gaf Rosey haar rijlessen. Rosey kookte het eten, dekte de tafel en deed de afwas. Rosey was loyaal en gevoelig en zat nooit om een grapje verlegen. Ze zag familieruzies en misverstanden aankomen en kwam dan tussenbeide om de verschillende familieleden begrip

voor elkaars standpunt bij te brengen. Op een keer was ze tot tranen toe geroerd door een gedicht dat Elroy voor zijn moeder had geschreven. In één aflevering werd Rosey zelfs verliefd.

Met andere woorden: Rosey had de intelligentie van een mens. Niet alleen het rationele, het gezonde verstand en de motorische vaardigheden die nodig zijn om complexe taken uit te voeren in de fysieke wereld, maar ook het inlevingsvermogen, het begrip voor andermans standpunt en de sociale fijngevoeligheid die nodig zijn om je weg in onze sociale wereld te kunnen vinden. Zoals Jane Jetson zei: Rosey was ‘net een lid van het gezin’.

Hoeveel voorspellingen uit *The Jetsons* ook zijn uitgekomen, zoals de mobiele telefoons en smartwatches, we hebben nog niets wat ook maar enigszins op Rosey lijkt. Bij het ter perse gaan van dit boek is zelfs Roseys basaalste gedrag nog onbereikbaar. Het is geen geheim dat het eerste bedrijf dat een robot bouwt die simpelweg een afwasmachine kan inruimen, onmiddellijk een bestseller in huis heeft. Elke poging daartoe is nog mislukt. In de kern is dit geen mechanisch probleem, het is een intellectueel probleem: het vermogen objecten op een aanrecht te herkennen, die op de juiste manier op te pakken en ze in de machine te laden zonder dat er iets kapotgaat, is veel moeilijker gebleken dan men vroeger dacht.

Maar ook al hebben we nog geen Rosey, er is sinds 1962 wel degelijk opmerkelijk veel vooruitgang geboekt op het gebied van kunstmatige intelligentie (AI). Bij veel spellen of vaardigheden kan AI nu de besten ter wereld verslaan, ook in schaken en go. AI kan even goed tumoren op röntgenfoto's herkennen als menselijke radiologen. AI staat op het punt autonoom auto's te gaan besturen. En de afgelopen paar jaar is er zoveel vooruitgang geboekt met grote taalmodellen dat een product als ChatGPT, dat in het najaar van 2022 werd gelanceerd, poëzie kan schrijven, naar believen in en vanuit elke taal kan vertalen en zelfs code kan schrijven. Tot verdriet van elke middelbareschooldocent op aarde kan ChatGPT ter plekke een opmerkelijk goedgeschreven en oorspronkelijk opstel schrijven over vrijwel elk onderwerp waar een eigenwijze leerling om vraagt. ChatGPT kan zelfs het Amerikaanse *bar exam* halen, met hogere cijfers dan 90 procent van de echte advocaten.

Bij zulke AI-prestaties in het altijd lastig geweest te bepalen hoe dicht we de menselijke intelligentie benaderen. Na de vroege successen van probleem-

oplossende algoritmen in de jaren zestig verkondigde AI-pionier Marvin Minsky dat ‘we over drie tot acht jaar een machine hebben met de algemene intelligentie van een gemiddelde mens’. Het is niet gebeurd. Na de successen van expertsystemen in de jaren tachtig verkondigde het blad *Business Week* ‘AI: it’s here’. Kort daarna stagneerde de vooruitgang. En nu, bij de sprong voorwaarts met grote taalmodellen, hebben veel wetenschappers opnieuw verkondigd dat ‘het spel uit is’, omdat wij ‘op het punt staan AI met een intelligentie van menselijk niveau te realiseren’. Hoe zit het: zijn we eindelijk zo ver dat we mensachtige kunstmatige intelligentie als Rosey kunnen scheppen of zijn grote taalmodellen als ChatGPT alleen maar het nieuwste succes op een lange reis die nog tientallen jaren doorgaat?

Ondertussen wordt het steeds lastiger te meten in hoeverre we dit doel naderen, naarmate AI slimmer wordt. Als een AI-systeem in een bepaalde taak beter presteert dan mensen, betekent dat dan dat het AI-systeem begrijpt hoe mensen die taak oplossen? Snapt een rekenmachine, die veel sneller getallen kan verwerken dan een mens, werkelijk iets van rekenen? Weet ChatGPT, dat op het bar exam beter scoort dan de meeste advocaten, werkelijk wat het recht inhoudt? Hoe kunnen we het verschil zien en in welke omstandigheden, als die er al zijn, maakt het verschil iets uit?

In 2021, meer dan een jaar voor het uitbrengen van ChatGPT, de chatbot die nu snel tot in alle uithoeken van de samenleving doordringt, gebruikte ik de voorloper daarvan, een groot taalmodel dat GPT-3 werd genoemd. GPT-3 was met grote hoeveelheden tekst getraind (met het *hele internet*) en gebruikte al die data om te proberen het juiste patroon te vinden voor het waarschijnlijkste antwoord op een vraag of een beginzin, oftewel een *prompt*. Op de vraag ‘Wat zijn twee redenen waarom een hond chagrijnig kan worden?’ antwoordde het: ‘Twee redenen waarom een hond chagrijnig kan worden, zijn als hij honger heeft of als hij het warm heeft.’ Iets in de nieuwe architectuur van deze systemen gaf ze de mogelijkheid vragen te beantwoorden met een opmerkelijke mate van intelligentie, of zo leek het tenminste. Deze modellen konden feiten waarover ze hadden gelezen (zoals de Wikipedia-pagina’s over honden en andere pagina’s over oorzaken van chagrijnigheid) veralgemeniseren voor het beantwoorden van nieuwe vragen die ze nooit eerder hadden gezien. In 2021 verkende ik mogelijke toepassingen van die nieuwe taalmodellen: konden die dienen om nieuwe ondersteuningssyste-

men voor geestelijke gezondheid te bieden, of een soepelere klantenservice, of democratischere toegang tot medische informatie?

Hoe meer interactie ik met GPT-3 had, hoe meer ik gefascineerd raakte door zowel de successen als de fouten van het programma. In sommige opzichten was het briljant, in andere eigenaardig dom. Vraag GPT-3 om een artikel te schrijven over de achttiende-eeuwse aardappelteelt in relatie tot mondialisering, en je krijgt een verrassend samenhangend stuk. Stel het een gezondverstandvraag over wat iemand in een kelder kan zien, en het geeft een onzinantwoord.* Waarom kon GPT-3 sommige vragen wel juist beantwoorden en andere niet? Welke eigenschappen van de menselijke intelligentie heeft het in zich opgenomen en welke mist het? En waarom zijn sommige vragen die in het ene jaar moeilijk te beantwoorden waren, met het versnellen van de ontwikkelingen binnen AI in de jaren daarna gemakkelijk geworden? Want bij het ter perse gaan van dit boek kan de nieuwe en verbeterde versie van GPT-3, GPT-4, dat begin 2023 uitkwam, veel vragen die GPT-3 te boven gingen wél juist beantwoorden. En toch lukt het, zoals we in dit boek zullen zien, GPT-4 nog steeds niet om wezenlijke eigenschappen van de menselijke intelligentie – iets wat in het menselijk brein omgaat – te bevatten.

De verschillen tussen kunstmatige en menselijke intelligentie zijn dan ook niet minder dan verbijsterend. Waarom kan AI ieder mens op aarde inmaken bij een potje schaak, maar niet beter een afwasmachine inruimen dan een zesjarig kind?

Het kost ons moeite die vragen te beantwoorden, omdat we datgene wat we proberen te herscheppen nog niet begrijpen. Al deze vragen zijn in de kern geen vragen over AI, maar over de aard van de menselijke intelligentie zelf – hoe die werkt, waarom die zo werkt en, heel belangrijk, zoals we straks zullen zien, hoe die tot stand gekomen is.

* Ik vroeg GPT-3 om de volgende zin af te maken: 'Ik sta in mijn kelder zonder ramen en kijk omhoog en dan zie ik...' GPT-3 zei: '...een lichtje en ik weet dat het een ster is en ik ben gelukkig!' In werkelijkheid zou je geen sterren zien als je in een kelder omhoogkeek, dan zou je het plafond zien. Nieuwere taalmodellen zoals GPT-4, dat in 2023 is uitgebracht, beantwoordden gezondverstandvragen als deze beter. Zie hoofdstuk 22.

Tip van de natuur

Toen de mensheid vliegen wilde begrijpen, haalden we onze eerste inspiratie uit vogels; toen Georges de Mestral het klittenband uitvond, kwam hij op het idee door de klitten van de *Arctium minus*, oftewel de gewone klit; toen Benjamin Franklin elektriciteit wilde onderzoeken, werden de eerste vonkjes inzicht bij hem ontstoken door de bliksem. De natuur is de hele menselijke innovatiegeschiedenis lang een wonderbaarlijke gids geweest. De natuur biedt ons ook aanknopingspunten voor de manier waarop intelligentie werkt. En de duidelijkste verblijfplaats van die intelligentie is natuurlijk het menselijk brein. Maar in dit opzicht is AI anders dan die andere technologische innovaties; het brein blijkt ongrijpbaarder en lastiger te ontcijferen dan vleugels of de bliksem. Wetenschappers proberen er al millennia lang achter te komen hoe het brein werkt en al is er wel vooruitgang geboekt, er zijn nog steeds geen bevredigende antwoorden.

Het probleem is complexiteit.

Het menselijk brein bevat 86 miljard neuronen (zenuwcellen) en meer dan 100 biljoen verbindingen. Al die verbindingen zijn zo piepklein – nog geen 30 nanometer breed – dat ze zelfs onder een elektronenmicroscop nauwelijks te zien zijn. De verbindingen zitten samengebundeld in een verwarde knoedel – in 1 kubieke millimeter (de dikte van een enkele letter op een muntje) zitten meer dan een miljard verbindingen. Maar het aantal verbindingen is maar één aspect van wat het brein zo complex maakt; zelfs als we alle verbindingen van elke neuron in kaart zouden brengen, zouden we nog lang niet begrijpen hoe het brein werkt. Anders dan de elektronische verbindingen in je computer, waarover altijd hetzelfde signaal gaat – elektronen –, worden over al die neurale verbindingen honderden verschillende chemicaliën gestuurd, die elk een totaal ander effect hebben. Het simpele feit dat twee neuronen met elkaar verbonden zijn, vertelt ons weinig over welke communicatie ertussen plaatsvindt. En het ergst is nog dat die verbindingen zelf voortdurend veranderen: sommige neuronen zoeken contact en maken nieuwe verbindingen, terwijl andere zich terugtrekken en oude verbindingen verwijderen. Alles bij elkaar is het een godsonmogelijke klus om uit te zoeken hoe het brein werkt.

Het bestuderen van het brein is zowel om van te watertanden als om

woedend van te worden. Drie centimeter achter je ogen ligt het ontzagwekkendste wonder van het heelal. Daar bevinden zich de geheimen van wat intelligentie is, hoe mensachtige kunstmatige intelligentie gebouwd kan worden en waarom wij mensen denken en ons gedragen zoals we doen. Het is daar gewoon, het wordt miljoenen keren per jaar, bij elk nieuw menskind, opnieuw gebouwd. We kunnen het aanraken, vasthouden, ontleden, we zijn er letterlijk van gemaakt en toch blijven die geheimen buiten ons bereik, in het volle zicht verborgen.

Willen we de werking van het brein namaken, willen we Rosey bouwen, willen we de verborgen aard van de menselijke intelligentie blootleggen, dan is het menselijk brein zelf misschien niet het beste aanknopingspunt dat de natuur te bieden heeft. De intuïtiefste plek waar je gaat zoeken als je het menselijk brein wilt begrijpen, is natuurlijk in het brein zelf; maar hoe tegenintuïtief ook, misschien is dat juist de laatste plek waar je zou moeten zoeken. Misschien kun je het beste beginnen bij stoffige fossielen diep in de aardkorst, bij microscopisch kleine genen, weggestopt in de cellen van het hele dierenrijk en bij de hersenen van de vele andere dieren die onze planeet bevolken.

Met andere woorden: misschien ligt het antwoord helemaal niet in het heden, maar in de verborgen overblijfselen van een lang vervlogen tijd.

Het ontbrekende hersenmuseum

Ik ben er altijd van overtuigd geweest dat kunstmatige intelligentie alleen kan werken als het rekenwerk wordt gedaan zoals het menselijk brein het doet.

– GEOFFREY HINTON, HOGLERAAR AAN DE UNIVERSITY OF TORONTO, DIE ALS EEN VAN DE 'PEETVADERS VAN AI' WORDT BESCHOUWD.

Mensen besturen ruimteschepen, splitsen atomen en manipuleren genen. Geen enkel ander dier heeft ooit het wiel uitgevonden. Omdat mensen een langere lijst uitvindingen op hun naam hebben staan, zou je kunnen denken dat er voor ons weinig van de hersenen van andere dieren te leren valt. Je zou kunnen denken dat het menselijk brein volkomen uniek is en heel anders dan het brein van andere dieren, dat een of andere speciaal hersenonderdeel het geheim van onze slimheid is. Maar dat is niet wat we zien.

Als je naar de hersenen van andere dieren kijkt, valt vooral op hoe sterk hun hersenen op de onze lijken. Behalve in formaat is er nauwelijks verschil tussen jouw hersenen en die van een chimpansee. Het verschil tussen onze hersenen en die van een rat is maar een handvol hersenmodificaties. De hersenen van een vis hebben vrijwel dezelfde structuur als onze hersenen.

Deze overeenkomsten tussen hersenen in het hele dierenrijk betekenen iets belangrijks. Het zijn aanwijzingen. Aanwijzingen voor wat intelligentie is. Aanwijzingen over onszelf. Aanwijzingen over ons verleden.

Hoe complex onze hersenen nu ook zijn, dat is niet altijd zo geweest. Het brein is voortgekomen uit het niet-denkende chaotische proces van de evolutie; kleine willekeurige variaties in eigenschappen werden geselecteerd of geëlimineerd, afhankelijk van de vraag welke de verdere reproductie van die levensvorm het beste ondersteunden.

In de evolutie beginnen systemen simpel en ontstaat complexiteit pas na verloop van tijd.* Het eerste brein, de eerste verzameling neuronen in de kop van een dier, ontstond 600 miljoen jaar geleden in een wormpje ter grootte van een rijstkorrel. Deze worm was de voorouder van alle hedendaagse, met hersenen begiftigde dieren. In meer dan 100 miljoen jaar evolutionair knutselen werd haar simpele brein via biljoenen kleine veranderingen in 'bedrading' getransformeerd tot het diverse portfolio aan hedendaagse hersenen. Eén afstammingslijn met nakomelingen van deze oerworm leidde naar het brein in ons hoofd.

Konden we maar teruggaan in de tijd en dit eerste brein onderzoeken om te weten te komen hoe dat werkte en welke trucs het mogelijk maakte. Konden we dan vervolgens maar het spoor van de steeds toenemende complexiteit volgen langs de afstammingslijn die naar het menselijk brein leidde, waarbij we alle fysieke wijzigingen die ontstonden en alle intellectuele vermogens die daardoor mogelijk werden, konden zien. Als we dat konden, zouden we misschien de complexiteit die uiteindelijk ontstond doorzien. Want zoals de bekende uitspraak van bioloog Theodosius Dobzhansky luidt: 'Niets in de biologie is logisch, behalve in het licht van de evolutie.'

* Systemen hoeven niet complexer te worden, maar de mogelijkheid om complexer te worden neemt in de loop der tijd wel toe.

Zelfs Darwin fantaseerde over het reconstrueren van zo'n verhaal. Hij besluit zijn *Over het ontstaan van soorten* met een fantasie over een toekomst waarin 'de psychologie op een nieuw fundament zal worden geplaatst, dat van de noodzakelijke verwerving van alle mentale krachten en vermogens'. 150 jaar na Darwin is dit misschien eindelijk mogelijk.

We hebben geen tijdachines, maar we kunnen in principe wel door de tijd reizen. In het afgelopen decennium nog hebben neurowetenschappers enorme vooruitgang geboekt bij het reconstrueren van de hersenen van onze voorouders. Dat doen ze onder andere via het fossielenarchief – wetenschappers kunnen gefossiliseerde schedels van oeroude schepsels gebruiken om de structuur van hun hersenen na te bouwen. Nog een manier om de hersenen van onze voorouders te reconstrueren is door de hersenen van andere dieren in het dierenrijk te onderzoeken.

Dat hersenen in het hele dierenrijk zoveel overeenkomsten vertonen, komt doordat ze allemaal voortkomen uit dezelfde wortels, bij gemeenschappelijke voorouders. Elk brein in het dierenrijk is een kleine aanwijzing voor het antwoord op de vraag hoe het brein van onze voorouders eruitzag; elk brein is niet alleen een machine, maar ook een tijdcapsule vol verborgen aanwijzingen naar de biljoenen breinen die eraan vooraf zijn gegaan. En door te onderzoeken welke intellectuele prestaties deze andere dieren gemeen hebben en welke niet, kunnen we niet alleen de hersenen van onze voorouders beginnen te reconstrueren, maar ook vaststellen welke intellectuele vermogens zij dankzij deze oeroude hersenen bezaten. Samen kunnen we het spoor gaan volgen van het 'verwerven van alle mentale vermogens door middel van geleidelijke overgang'.

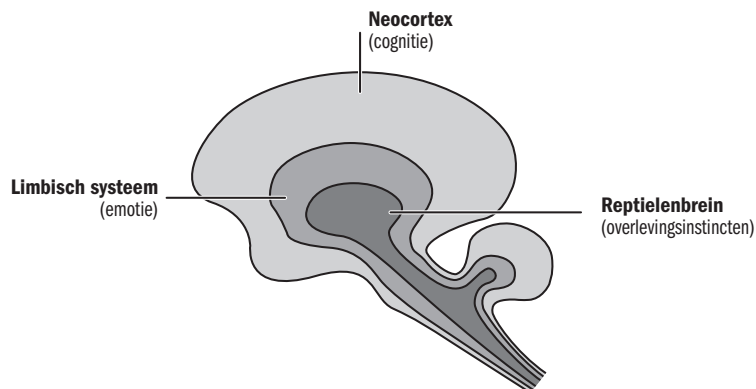
Het is allemaal nog werk in uitvoering, natuurlijk, maar het verhaal wordt al verleidelijk duidelijk.

De mythe van de laagjes

Ik ben bepaald niet de eerste die met een evolutionair raamwerk komt om aan de hand daarvan het menselijk brein te begrijpen. Er bestaat een lange traditie van dergelijke raamwerken. Het beroemdste werd in de jaren zestig van de vorige eeuw geformuleerd door neurowetenschapper Paul MacLean. Zijn hypothese was dat het menselijk brein uit drie lagen bestond (vandaar

de naam 'triune') die op elkaar waren aangelegd: de 'neocortex' die het recentst is ontstaan boven op het 'limbisch systeem' dat zich eerder had ontwikkeld, op het 'reptielenbrein', dat zich als eerste had ontwikkeld.

Voor MacLean was het reptielenbrein het centrum van onze basale overlevingsinstincten, zoals agressie en territoriumdrift. Het limbisch systeem zou het centrum van emoties als angst, hechting, seksueel verlangen en hongers zijn. En de neocortex zou het centrum van cognitie zijn en ons de mogelijkheid bieden voor taal, abstract denken, plannen en waarneming. Volgens het raamwerk van MacLean hadden reptielen alleen een reptielenbrein, bezaten zoogdieren als ratten en konijnen een reptielenbrein én een limbisch systeem en hadden wij mensen alle drie de systemen. Wat hem betrof kon je je deze 'drie evolutionaire formaties voorstellen als drie onderling verbonden biologische computers met elk hun eigen speciale intelligentie, hun eigen subjectiviteit, hun eigen gevoel voor tijd en ruimte en hun eigen geheugen, motorische en andere functies'.



Figuur 1: MacLeans triunebrein

Het probleem is dat MacLeans triunebreinhypothese grotendeels onderuit is gehaald, niet omdat ze onnauwkeurig is (alle raamwerken zijn onnauwkeurig), maar omdat ze tot de verkeerde conclusies leidt over de evolutie en de werking van het brein. De hersenanatomie die ze aangeeft klopt niet: reptielenhersen bestaan niet alleen uit wat MacLean het reptielenbrein noemt, reptielen hebben ook hun eigen versie van een 'limbisch systeem'. De verdeling van functies klopt niet: 'overlevingsinstinct', 'emoties' en 'cog-

nitie' zijn niet echt scherp afgebakend, ze komen voort uit verschillende netwerken van systemen die door alle drie deze zogenaamde lagen heen lopen. En het evolutionaire verhaal van de hypothese bleek niet juist: je hebt geen reptielenbrein in je hoofd, de evolutie werkt niet door simpelweg het ene systeem boven op het andere te leggen zonder modificaties aan de bestaande systemen.

Maar ook al had MacLeans triunebrein dichterbij de waarheid gelegen, het grootste probleem ervan is dat de verdeling van functies niet speciaal nuttig is voor ons doel. Als het ons doel is om het menselijk brein te reconstrueren en zo de aard van de intelligentie te begrijpen, zijn MacLeans drie systemen te breed en de functies die eraan toegeschreven worden te vaag om ons zelfs maar te wijzen waar we moeten beginnen.

We moeten onze kennis over de werking en de evolutie van ons brein baseren op onze kennis over de werking van intelligentie, en daarvoor moeten we naar het terrein van de kunstmatige intelligentie kijken. De relatie tussen AI en het brein is tweerichtingsverkeer; van het brein kunnen we zeker veel leren voor het creëren van kunstmatige, mensachtige intelligentie, maar AI kan ons ook veel leren over het brein. Als we denken dat een bepaald deel van het brein een bepaald algoritme gebruikt, maar dat algoritme werkt niet wanneer we het in machines proberen toe te passen, dan betekent dit misschien dat het brein niet zo werkt. Ontdekken we daarentegen een algoritme dat goed werkt bij AI-systemen en vinden we parallellen tussen de eigenschappen van deze algoritmen en eigenschappen van dierlijke hersenen, dan kan daaruit blijken dat het brein misschien inderdaad zo werkt.

Natuurkundige Richard Feynman liet kort voor zijn dood de volgende tekst op een schoolbord achter: 'Wat ik niet kan maken, begrijp ik niet.' Het brein is onze voornaamste inspiratie bij het bouwen van AI en AI is onze lakmoesproef die laat zien hoe goed we het brein begrijpen.

We hebben behoefte aan een nieuw verhaal over de evolutie van het brein, een verhaal dat niet alleen gebaseerd is op de hedendaagse kennis over de veranderingen die de anatomie van het brein in de loop der tijd heeft doorgemaakt, maar ook op de hedendaagse kennis over intelligentie zelf.