

OPEN CIVITAS

Essays

De ontmythologisering van menselijke intelligentie

Een filosofische analyse van kunstmatige intelligentie en haar effect op het denken over menselijke intelligentie, aan de hand van Hofstadter, Mitchell en Dennett

Dick H. Ahles





Colofon

© 2026 Dick H. Ahles

Uitgever: Open Civitas

Reeks: Essays

Datum: 11 oktober 2026

Versie: DA04

Over de auteur



Dick H. Ahles studeerde politicologie aan de Vrije Universiteit Amsterdam, met bijzondere aandacht voor onderzoeksmethoden en statistiek. Als onafhankelijk auteur onderzoekt hij civiliseringsprocessen, politieke overtuigingen, institutionele macht en maatschappelijke verandering. Hij publiceert via Open Civitas en woont en werkt in Enkhuizen.

Inhoudsopgave

Samenvatting.....	4
1 Inleiding.....	4
1.1 Vraagstelling en werkwijze.....	5
2 Waarschuwingen en gebruik van AI.....	5
3 Hofstadter en GEB.....	6
4 Mitchell als gids en gesprekspartner.....	7
5 Twee betekenissen van angst.....	8
6 De verbijstering bij Google.....	9
7 Formele eenvoud en onverwachte complexiteit.....	10
8 Wetenschappelijke verklaring en onttovering.....	11
9 De premisse van menselijke uitzonderlijkheid.....	11
10 Betekenis zonder begrip.....	12
11 Wat Hofstadters angst kan rechtvaardigen.....	14
12 Darwins merkwaardige omkering.....	15
13 Mitchell na ChatGPT.....	15
14 Begrip als geleidelijk vermogen.....	16
15 Praktisch risico en metafysische angst.....	17
16 Van vervanging naar vergroting.....	18
17 Conclusie.....	19
Literatuur.....	20

Samenvatting

Kunstmatige intelligentie, zo lijkt het, roept vaak twee verschillende soorten angst op bij mensen. De eerste betreft de angst voor AI-systemen die aan menselijke controle ontsnappen en schade veroorzaken, inclief de vrees dat die systemen zich tegen mensen zal keren. De tweede ontstaat wanneer machines vermogens vertonen die wij altijd als bewijs van menselijke uitzonderlijkheid hebben beschouwd. Dit essay onderzoekt vooral die tweede angst. De centrale stelling luidt dat AI de mens niet noodzakelijkerwijze ‘degradeert’, maar wel de ondeugdelijke argumenten voor de menselijke grootheid ondergraaft. En dat laatste is voor velen wel een gevaarlijk idee.

Trefwoorden: kunstmatige intelligentie; menselijke intelligentie; ontmythologisering; begrip; Douglas Hofstadter; Melanie Mitchell; Daniel Dennett

1 Inleiding

Het publieke debat over kunstmatige intelligentie wordt beheerst door verwachtingen die sterk uiteenlopen. Sommige auteurs stellen vooral de wenselijkheid van automatisering ter discussie. Christopher Roosen vraagt bijvoorbeeld welke betekenis het schrijven voor de menselijke schrijver heeft, ook wanneer een systeem overtuigende teksten kan produceren.¹ Abeba Birhane en Marek McGann betogen vanuit een enactivistisch perspectief dat de taalprestaties van grote taalmodellen niet gelijkstaan aan menselijke linguïstische handelingsbekwaamheid.² Binnen invloedrijke delen van de AI-wereld leeft daarentegen de verwachting dat AI zich binnen afzienbare tijd kan ontwikkelen tot artificial general intelligence (AGI) en zich vervolgens mogelijk verder ontwikkelt tot kunstmatige superintelligentie. Met AGI wordt doorgaans een systeem bedoeld dat niet slechts één nauw omschreven taak uitvoert, maar op uiteenlopende gebieden kan leren, redeneren en problemen oplossen. Superintelligentie duidt vervolgens op een systeem dat de menselijke intellectuele vermogens op vrijwel alle relevante gebieden overtreft. Over de precieze betekenis van deze begrippen bestaat geen overeenstemming en evenmin staat vast dat de huidige technologie langs een geleidelijke ontwikkelingslijn tot zulke systemen zal leiden.³

Voor anderen is de mogelijke ontwikkeling naar AGI of superintelligentie vooral beangstigend omdat autonome systemen aan menselijke controle zouden kunnen ontsnappen of tegen menselijke belangen zouden kunnen handelen. Zulke verwachtingen krijgen soms een uitgesproken dramatische vorm in de publieke berichtgeving.⁴ Daarnaast bestaan zorgen over ingrijpende economische gevolgen van AI. Die leiden tot oproepen tot onderzoek, maatschappelijke

¹ Christopher Roosen, “It Doesn’t Matter If Artificial Intelligence Can Write a Novel—It’s the Wrong Thing to Automate” (2 augustus 2023), online-essay: <https://www.christopherroosen.com/blog/2023/8/2/story-engine-word-machines-writing-automation>. Roosen bespreekt de wenselijkheid van automatisering, niet alleen de kwaliteit van de geproduceerde tekst.

² Abeba Birhane en Marek McGann, “Large Models of What? Mistaking Engineering Achievements for Human Linguistic Agency”, *Language Sciences* 106 (2024), artikel 101672. Openbare auteursversie: <https://doi.org/10.48550/arXiv.2407.08790>.

³ Katja Grace e.a., “Thousands of AI Authors on the Future of AI”, *Journal of Artificial Intelligence Research* 84 (2025), <https://doi.org/10.1613/JAIR.1.19087>. Het betreft geaggregeerde inschattingen van 2.778 onderzoekers, geen vastgestelde ontwikkelingslijn. Zie ook Tim Genewein e.a., “From AGI to ASI” (2026), <https://doi.org/10.48550/arXiv.2606.12683>, over vier mogelijke ontwikkelingsroutes en hun onzekerheden.

⁴ Frank Bergman, “Top Investor Sounds Alarm: AI Elites Are Trying to Create a ‘God-Like’ Intelligence to Rule Humanity”, *Slay News* (7 juni 2026): <https://slaynews.com/top-investor-sounds-alarm-ai-elites-trying-create-god-like-intelligence-rule-humanity/>. Dit nieuwsbericht is een voorbeeld van de publieke voorstelling van AI; het is geen technisch bewijs voor de beschreven intenties of risico’s.

voorbereiding en publiek beleid.⁵ Een voorstelling van een vijandige machine en een economische oproep tot voorbereiding zijn verschillende soorten argumenten en moeten afzonderlijk worden beoordeeld.

De onzekerheid over de technische ontwikkeling verklaart slechts een deel van de intensiteit van het debat. AI raakt ook aan de wijze waarop mensen zichzelf begrijpen. Vermogens als spreken, redeneren, plannen, programmeren, componeren en het leggen van analogieën golden lange tijd als eigenschappen die een scherpe grens trokken tussen mensen en machines. Wanneer een kunstmatig systeem op deze gebieden overtuigende prestaties levert, ontstaat daarom meer dan een discussie over techniek. Er komt een antropologische en filosofische vraag in beeld: wat blijft er over van de gedachte dat menselijke intelligentie een unieke, principieel onnavolgbare eigenschap is?

Aan de andere kant zijn er onderzoekers en ontwikkelaars die het belang en het nut van verdere ontwikkeling van kunstmatige intelligentie benadrukken. Sinds de Tweede Wereldoorlog is langs uiteenlopende lijnen gewerkt aan systemen die aspecten van menselijke intelligentie modelleren of taken van mensen overnemen. Mitchell geeft van die geschiedenis een overzicht.⁶ Het huidige onderzoek vindt mede plaats in een internationale concurrentiestrijd tussen bedrijven. De deskundigheid van de betrokken onderzoekers verdient erkenning, maar vormt op zichzelf geen garantie voor de juistheid van iedere verwachting over de verdere ontwikkeling.

Dit essay beoordeelt niet in algemene zin of AI nuttig of foutloos is, welke gevaren aan haar gebruik verbonden zijn of welke maatregelen nodig zijn. Het onderzoekt hoe het bestaan van AI-systemen en de verwachtingen over hun vermogens onze vooronderstellingen over menselijke intelligentie ter discussie stellen. Die verschuiving duid ik aan als ontmythologisering.

1.1 Vraagstelling en werkwijze

De centrale vraag is welke conclusies over menselijke intelligentie en waardigheid gerechtvaardigd zijn wanneer kunstmatige systemen vergelijkbare cognitieve prestaties leveren. Dit essay is een filosofisch en argumentatief onderzoek. Ik lees Hofstadter, Mitchell en Dennett in relatie tot die vraag en onderscheid de beoordeling van een prestatie van de beoordeling van het proces of de producent. Persoonlijke ervaringen dienen als illustraties van een verschuiving in denkschema's. Zij leveren geen empirisch bewijs voor de vermogens van AI of voor de motieven van haar critici.

2 Waarschuwingen en gebruik van AI

In dit essay maak ik onderscheid tussen twee vormen van waarschuwingen voor AI. De eerste is praktisch en politiek. Zij betreft schade die kan ontstaan door verkeerde doelen, gebrekkige controle, menselijk misbruik of een toenemende maatschappelijke afhankelijkheid van ondoorzichtige systemen. De tweede vorm is conceptueel en existentieel. Zij ontstaat wanneer AI menselijke intelligentie ontmythologiseert door te laten zien dat althans sommige prestaties die wij aan innerlijke diepgang verbonden, ook uit een niet-menselijk proces kunnen voortkomen.

⁵ Erik Brynjolfsson, Ajay Agrawal, Anton Korinek en Tom Cunningham (organisatoren), "We Must Act Now: A Statement on AI's Transformation of the Economy" (juli 2026); <https://www.wemustactnow.ai/>. De verklaring vraagt om onderzoek en voorbereiding op economische veranderingen en bepleit dat AI menselijke vermogens aanvult.

⁶ Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (With a New Preface) (New York: Picador, 2025), hoofdstuk 1, "The Roots of Artificial Intelligence". Eerste editie: 2019.

Beide vormen kunnen tegelijk optreden, maar zij vereisen verschillende argumenten en leiden tot verschillende vragen.

Douglas Hofstadters uitspraak, zoals beschreven door Melanie Mitchell, dat hij terrified is, vormt het uitgangspunt van de analyse. Zijn angst is opmerkelijk omdat juist zijn werk een van de krachtigste intellectuele voorbereidingen biedt op het idee dat intelligentie uit de wisselwerking van niet-intelligente onderdelen kan ontstaan. Melanie Mitchell, zijn voormalige promovenda en medewerker, gebruikt Hofstadters reactie als opening van haar boek over AI. Haar nieuwe voorwoord uit 2025 maakt de spanning nog scherper. Zij erkent dat recente systemen vermogens hebben ontwikkeld die zij enkele jaren eerder nog als belangrijke ontbrekende voorwaarden voor werkelijke intelligentie beschouwde, maar blijft twijfelen of deze systemen werkelijk begrijpen.⁷

Mijn vraag is niet of Hofstadters emotie geoorloofd is. Wetenschappelijke en technische ontdekkingen kunnen vreugde, ontroering, afkeer en angst oproepen. De relevante vraag is welke conclusie uit die emotie mag worden getrokken. Ik ben van mening dat de praktische risico's van AI serieus moeten worden onderzocht, maar dat de ontmythologisering van menselijke intelligentie op zichzelf geen reden tot angst is. Zij dwingt ons wel om menselijke waardigheid op een steviger grondslag te plaatsen dan het veronderstelde monopolie op bepaalde cognitieve prestaties. De gedachte dat een computer met chips en een programma prestaties kan leveren die gelijkwaardig zijn aan, of zelfs beter dan, die van een menselijk brein, kan moeilijk voorstelbaar en beangstigend zijn.

Ondanks de waarschuwingen is het gebruik van AI in korte tijd sterk toegenomen. Een in januari 2026 gepubliceerd internationaal onderzoek van Google en Ipsos rapporteerde dat gemiddeld 66 procent van de ondervraagde volwassenen in 21 landen in de voorafgaande twaalf maanden een AI-toepassing had gebruikt. Het veldwerk vond plaats in september en oktober 2025. De AI Index 2026 meldt dat in 2025 88 procent van de onderzochte organisaties AI gebruikte en 70 procent generatieve AI in ten minste één bedrijfsfunctie. OpenAI meldde in augustus 2026 meer dan één miljard ChatGPT-gebruikers.⁸ Deze cijfers zijn door verschillen in definities, perioden, steekproeven en onderzoeksmethoden niet rechtstreeks vergelijkbaar. Zij wijzen wel op brede toepassing van de technologie. Gebruik zegt op zichzelf niets over begrip, betrouwbaarheid of de mate van intelligentie van de gebruikte systemen.

3 Hofstadter en GEB

Douglas Hofstadter behoort niet tot de technische grondleggers van de neurale netwerken waarop de huidige generatieve AI berust. Toch geldt hij binnen het denken over kunstmatige intelligentie als een van de invloedrijkste en oorspronkelijkste auteurs, bijna als een geestelijk peetvader van het vakgebied. Die positie dankt hij vooral aan *Gödel, Escher, Bach: An Eternal Golden Braid*, doorgaans afgekort tot GEB. Het boek verscheen in 1979 en ontving in 1980 de Pulitzerprijs voor algemene non-fictie. Het is geen leerboek over programmeren en evenmin een conventionele filosofische verhandeling. Hofstadter verbindt er wiskundige logica, muziek,

⁷ Mitchell, *Artificial Intelligence* (2025), nieuw voorwoord.

⁸ Google en Ipsos, *Google/Ipsos Multi-Country AI Survey 2026* (15 januari 2026), <https://www.ipsos.com/en-us/google-ipsos-multi-country-ai-survey-2026>, met veldwerk van 22 september tot 10 oktober 2025; Stanford HAI, *The 2026 AI Index Report, "Economy"*, <https://hai.stanford.edu/ai-index/2026-ai-index-report/economy>, over organisatiegebruik in 2025; OpenAI, "From Asking to Doing: How the World Is Putting ChatGPT to Work" (6 augustus 2026), <https://openai.com/index/how-the-world-is-putting-chatgpt-to-work/>. Geraadpleegd op 11 oktober 2026.

beeldende kunst, taal, cognitiewetenschap en speelse dialogen in een onderzoek naar betekenis, zelfverwijzing en bewustzijn.

De centrale vraag van GEB is hoe intelligentie, betekenis en een besef van het eigen ik kunnen ontstaan uit processen die deze eigenschappen afzonderlijk niet bezitten. Hofstadter bespreekt formele systemen, recursie en zelfverwijzing en gebruikt de onvolledigheidsstellingen van Gödel om te laten zien dat de verhouding tussen een systeem en uitspraken over dat systeem ingewikkelder is dan een eenvoudige verzameling regels doet vermoeden. Zijn analyses hebben geen rechtstreeks technisch recept voor hedendaagse AI opgeleverd. Hun betekenis ligt in het denkmodel: eigenschappen van een georganiseerd geheel hoeven niet als afzonderlijke eigenschappen in de onderdelen aanwezig te zijn.

Het MU-systeem waarmee GEB begint, maakt dit denkmodel tastbaar. Een beperkt aantal regels bepaalt welke tekenreeksen uit andere tekenreeksen mogen worden afgeleid. Iedere afzonderlijke bewerking is eenvoudig en controleerbaar, terwijl de verzameling mogelijke afleidingen voor een menselijke waarnemer al snel ingewikkeld wordt. Het voorbeeld introduceert bovendien het verschil tussen handelen binnen een formeel systeem en redeneren over het systeem als geheel. Dat verschil loopt als een rode draad door GEB en wordt uiteindelijk verbonden met de vraag hoe een brein uitspraken over zichzelf kan vormen.⁹

Toen ik GEB kort na zijn verschijning las, veranderde het boek mijn manier van denken. Het liet mij zien dat onverwachte complexiteit geen geheimzinnige kracht vereist die van buitenaf aan een systeem wordt toegevoegd. Zij kan ontstaan uit de opeenstapeling en wisselwerking van op zichzelf heel veel eenvoudige processen. Daarmee bood Hofstadter ook een principieel antwoord op de vraag of machines intelligent zouden kunnen worden. Als menselijke intelligentie zelf voortkomt uit de organisatie van niet-intelligente bestanddelen (neuronen), bestaat er geen logische reden om intelligentie uitsluitend aan biologisch materiaal voor te behouden. Dat was destijds geen bewijs, en is ook nu nog geen bewijs dat iedere geschikte organisatie technisch realiseerbaar is, maar het neemt wel een principiële blokkade weg.

4 Mitchell als gids en gesprekspartner

Melanie Mitchell staat intellectueel dicht bij Hofstadter. Zij promoveerde bij hem en werkte met hem aan Copycat, een programma dat menselijke analogievorming probeerde te modelleren. Die achtergrond is van belang voor de toon van *Artificial Intelligence: A Guide for Thinking Humans*. Mitchell schrijft niet als een buitenstaander die een tegenstander bestrijdt. Zij deelt Hofstadters belangstelling voor abstractie, analogie en de vraag hoe betekenis ontstaat, maar beoordeelt de prestaties van AI met een consequente scepsis tegenover voortijdige claims over menselijke of bovenmenselijke intelligentie.

De eerste editie van haar boek verscheen in 2019, in een periode waarin diepe neurale netwerken grote vooruitgang boekten in beeldherkenning, taalverwerking, vertaling en spelstrategieën. Mitchell erkende die vooruitgang, maar benadrukte de broosheid van de systemen. Kleine veranderingen in een afbeelding konden tot absurde classificaties leiden. Taalprogramma's produceerden vloeiende zinnen zonder betrouwbare beheersing van context, dubbelzinnigheid en alledaagse kennis. Volgens haar werd de vooruitgang van AI overschat en de complexiteit van menselijke intelligentie onderschat.

⁹ Douglas R. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid* (New York: Basic Books, 1979), de behandeling van de MU-puzzel en het MIU-systeem.

In de editie met het nieuwe voorwoord uit 2025 kijkt Mitchell opnieuw naar dat oordeel. De verschijning van ChatGPT en andere grote taalmodellen heeft voorbeelden die zij eerder als duidelijke grenzen presenteerde, minder overtuigend gemaakt. Zij verbergt dat niet. Juist haar bereidheid de nieuwe prestaties serieus te nemen, maakt haar boek tot een geschikte leidraad voor dit essay. Haar positie bevindt zich tussen ongeclausuleerd enthousiasme en de eenvoudige ontkenning dat er iets wezenlijks is veranderd.

5 Twee betekenissen van angst

De eerste betekenis van angst voor AI betreft het verlies van controle. Een voldoende autonoom systeem zou doelen kunnen nastreven die niet overeenkomen met menselijke belangen. Ook zonder bewustzijn, haat of een verlangen naar macht kan een systeem schade veroorzaken wanneer het een verkeerd geformuleerd doel zeer doeltreffend uitvoert. Daarnaast kunnen mensen AI doelbewust gebruiken voor surveillance, manipulatie, cyberaanvallen of wapens. Nog waarschijnlijker is dat organisaties systemen inzetten waarvan zij de beperkingen onvoldoende begrijpen en hun beslissingen vervolgens een gezag toekennen dat zij niet verdienen.

In de populaire voorstelling krijgt dit controleprobleem vaak een menselijker gedaante. De machine wordt een actor die wil overleven, macht wil verwerven en zich uiteindelijk tegen de mens keert. Deze voorstelling sluit gemakkelijk aan bij sciencefiction en maakt van een mogelijk technisch en institutioneel probleem een strijd tussen twee soorten wezens. Daarmee wordt ook de gedachte versterkt dat een autonome AI bijna noodzakelijk vijandig zal zijn en nauwelijks zelfstandig behulpzaam zou kunnen handelen. Voor die noodzakelijkheid bestaat geen bewijs. Zowel behulpzaam als schadelijk gedrag hangt af van de architectuur, de doelen, de omgeving en de wijze waarop mensen het systeem inzetten.

De eerste angst leidt begrijpelijkerwijs tot de roep om onderzoek naar veiligheid, wettelijke grenzen en publiek toezicht. Over de gewenste maatregelen kan verschil van mening bestaan. Een verbod, vergunningenstelsel of moratorium kan grotere maatschappelijke kosten hebben dan een minder ingrijpende vorm van regulering. Maar de structuur van het vraagstuk is helder: welk concreet risico bestaat er, hoe waarschijnlijk is het, wie veroorzaakt het, welke schade kan optreden en welke maatregel vermindert die schade zonder grotere schade te creëren?

De tweede betekenis van angst heeft een ander object. Zij betreft niet in de eerste plaats wat AI met ons zou kunnen doen, maar wat het bestaan van AI over onszelf zou kunnen onthullen. Wanneer een machine kan schrijven, redeneren, componeren en nieuwe analogieën vormen, verliest de mens misschien eigenschappen waarop hij zijn uitzonderlijke positie baseerde. AI wordt dan een gevaarlijk idee omdat zij intelligentie ontdoet van haar status als een onverklaarbaar en exclusief menselijk vermogen. Deze angst leidt niet rechtstreeks tot een veiligheidsnorm. Zij leidt eerder tot pogingen de grens van echte intelligentie, begrip of creativiteit zo te definiëren dat de mens aan de veilige kant van de grens blijft.

De twee angsten worden in het publieke debat vaak vermengd. Daardoor kan een aantasting van menselijke status worden gepresenteerd als een veiligheidsprobleem en kan een concreet risico worden versterkt met beelden van een concurrerende kunstmatige soort. Voor een rationele beoordeling moeten zij analytisch worden gescheiden. Hofstadters terrified-reactie behoort vooral tot de tweede categorie. Zijn ontzetting ontstaat niet doordat hij een militair of politiek scenario beschrijft, maar doordat machines eigenschappen lijken te benaderen die hij met menselijke diepgang en creativiteit had verbonden.

6 De verbijstering bij Google

Waar ging de verbijstering van Hofstadter over? Mitchell begint de proloog van haar boek met een bijeenkomst bij Google. Hofstadter spreekt daar over de ontwikkeling van kunstmatige intelligentie en zegt zonder veel omhaal dat hij doodsbang is. Zijn formulering ‘I am terrified. Terrified.’ geeft de proloog haar titel en bepaalt de emotionele inzet van het boek. De uitspraak krijgt juist gewicht omdat zij afkomstig is van iemand die een groot deel van zijn intellectuele leven heeft besteed aan de vraag hoe geest en betekenis uit materiële processen kunnen ontstaan. Bovendien sloot zijn angst rechtstreeks aan bij een overtuiging die hij al veel eerder had geformuleerd. Aan het einde van *Gödel, Escher, Bach* stelde Hofstadter zichzelf in de paragraaf Ten Questions and Speculations tien vragen over de mogelijkheden van kunstmatige intelligentie. Een daarvan was of een computer ooit werkelijk mooie muziek zou kunnen schrijven. Zijn antwoord op die vraag luidde dat dit weliswaar denkbaar was, maar niet binnen afzienbare tijd: muziek was voor hem een taal van emoties, en zonder een gevoelsleven dat in complexiteit met het menselijke vergelijkbaar was, zou een programma volgens hem hooguit oppervlakkige imitaties kunnen produceren.¹⁰

Juist daarom vormde het latere werk van David Cope een confrontatie met een van Hofstadters eigen diepste aannames. Copes programma EMI (Experiments in Musical Intelligence) analyseerde bestaande composities en produceerde nieuwe muziek in de stijl van onder anderen Bach en Chopin. Luisteraars bleken zulke stukken niet altijd betrouwbaar van menselijke composities te kunnen onderscheiden. Voor Hofstadter raakte dit aan meer dan een geslaagde stijl-imitatie. Muziek, en vooral de muziek van componisten als Bach en Chopin, was voor hem verbonden met een menselijk leven, een gevoelswereld en een vermogen tot betekenisvolle expressie. Wanneer een programma vergelijkbare muziek kon voortbrengen, leek de mogelijkheid te ontstaan dat deze menselijke diepte voor het hoorbare resultaat niet noodzakelijk was.¹¹

Daarom interpreteer ik Hofstadters angst niet primair als vrees voor een agressieve machine. Zij betreft de gedachte dat intelligentie en creativiteit misschien gemakkelijker te produceren zijn dan hij had aangenomen en dat de menselijke geest uiteindelijk uit een verzameling betrekkelijk eenvoudige vermogens zou kunnen bestaan. Precies daar ontstaat mijn verbazing. De technische ontwikkeling lijkt geen weerlegging van GEB, maar een onverwachte bevestiging van een van zijn centrale mogelijkheden. Eenvoudige of lokaal beperkte processen kunnen op het niveau van het geheel resultaten voortbrengen die door geen afzonderlijk onderdeel zijn ontworpen of begrepen.

7 Formele eenvoud en onverwachte complexiteit

Het MU-systeem toont niet dat ieder eenvoudig regelsysteem vanzelf intelligentie voortbrengt. Evenmin bewijst het dat complexiteit noodzakelijk tot praktische onvoorspelbaarheid leidt. Het laat wel zien dat de eigenschappen van een formeel systeem niet volledig worden begrepen door alleen naar de eenvoud van iedere regel te kijken. De ruimte van mogelijke afleidingen, de terugkerende patronen en de grenzen van wat binnen het systeem kan worden bewezen, ontstaan uit de gezamenlijke werking van de regels.

¹⁰ Mitchell, *Artificial Intelligence* (2025), proloog, “Terrified”. Hofstadters reactie wordt hier gelezen via Mitchells weergave; de duiding ervan in dit essay is die van de auteur.

¹¹ David Cope, “Experiments in Musical Intelligence (EMI): Non-linear Linguistic-based Composition”, *Interface* 18, nr. 1–2 (1989): 117–139, <https://doi.org/10.1080/09298218908570541>.

Hetzelfde algemene inzicht keert terug in andere domeinen. Het weer volgt fysische processen, maar de wisselwerking van enorme aantallen variabelen en de gevoeligheid voor beginvoorwaarden begrenzen de voorspelbaarheid. Een stad groeit door afzonderlijke beslissingen van bewoners, ondernemingen, bestuurders en instellingen, terwijl het stedelijke patroon door niemand volledig wordt ontworpen. Als student politicologie aan de Vrije Universiteit ontwikkelde ik een dynamisch model van de groei van Amsterdam, wiskundig beschreven met 98 vrijheidsgraden.¹² Betrekkelijk eenvoudige gedragsregels en terugkoppelingen genereerden daarin ruimtelijke uitkomsten die vooraf niet vanzelfsprekend waren. De vele uiteenlopende varianten brachten mij zowel tot wanhoop als tot het inzicht dat lokale eenvoud geen eenvoudige voorspelling van het geheel garandeert.

Deze voorbeelden mogen niet zonder meer met elkaar worden gelijkgesteld. Een deductief formeel systeem, een chaotisch weersysteem, een sociaal model en een getraind neurale netwerk hebben verschillende structuren. De overeenkomst ligt op een hoger abstractieniveau. In elk geval kan de kennis van de lokale onderdelen onvoldoende zijn om het gedrag van het geheel intuïtief te voorspellen. De onverwachte uitkomst vereist daarom niet automatisch een bovennatuurlijke of onstoffelijke oorzaak. Zij kan een emergente eigenschap zijn van de organisatie en dynamiek van het systeem.

Voor kunstmatige intelligentie betekent dit dat het ontbreken van een expliciet door programmeurs ingevoerde regel geen bewijs vormt dat het systeem een vermogen niet kan bezitten. Moderne neurale netwerken worden niet regel voor regel geprogrammeerd om een foto te beschrijven of een analogie te formuleren. Zij worden getraind op grote hoeveelheden voorbeelden en ontwikkelen interne structuren die hun gedrag mogelijk maken. Dat gedrag kan verrassend zijn voor zowel gebruikers als ontwerpers. Verrassing bewijst geen bewustzijn, maar zij weerlegt wel de gedachte dat een computer uitsluitend kan doen wat een programmeur in de vorm van afzonderlijke instructies al volledig heeft voorzien.

Daaruit volgt geen voorspelling dat menselijke of bovenmenselijke intelligentie onvermijdelijk zal ontstaan. De organisatie kan ontoereikend blijken, training kan fundamentele beperkingen hebben en belangrijke menselijke vermogens kunnen afhankelijk zijn van lichamelijke en sociale processen die huidige systemen missen. De conclusie is bescheidener: wanneer intelligentie een emergent verschijnsel is, bestaat er geen principiële reden om uit de eenvoud van afzonderlijke bewerkingen af te leiden dat het geheel nooit intelligent gedrag kan ontwikkelen.

8 Wetenschappelijke verklaring en onttovering

Een wetenschappelijke ontdekking kan emotioneel ingrijpend zijn zonder dat de emotie deel uitmaakt van haar geldigheid. Newtons wetten veranderden het beeld van beweging en zwaartekracht. Darwins evolutietheorie veranderde het beeld van het ontstaan van soorten. Een theorie kan waar of onwaar, goed of slecht onderbouwd en ruim of beperkt toepasbaar zijn. De angst, vreugde of ontroering die zij oproept, kan ons iets vertellen over de betekenis die wij eraan toekennen, maar niet beslissen of de theorie houdbaar is.

Daarmee is wetenschap niet emotieloos. De ordening van het periodiek systeem kan schoonheid oproepen. Een onverwacht eenvoudig bewijs kan ontroeren doordat het uiteenlopende verschijnselen in één inzicht samenbrengt. Ik herinner mij hoe mijn wiskundeleraar integraalreke-

¹² Mijn scriptie is nooit gepubliceerd. Het model was een adaptatie van het systeem van Jay W. Forrester, *Urban Dynamics* (Cambridge, MA: MIT Press, 1969).

ning introduceerde. Hij had de gebruikelijke leerboeken uit zijn klas verbannen en de theorie van differentiaal- en integraalrekening zelf op stencils uitgeschreven. Op het bord tekende hij een grillige lijn met assen en vroeg hoe wij de oppervlakte onder die lijn konden bepalen.

De oppervlakte kon worden benaderd door de figuur in rechthoeken te verdelen waarvan de oppervlakte afzonderlijk exact kon worden berekend. Hoe smaller de rechthoeken, hoe meer oppervlakten moesten worden opgeteld en hoe dichter de som de gezochte oppervlakte naderde. Toen hij vroeg hoe wij de exacte oppervlakte konden vinden, bleef het stil in de klas. Zijn antwoord was: maak de rechthoekjes oneindig klein en tel dat bij elkaar op! Het idee verbijsterde mij en bracht tranen in mijn ogen. De formele regels over limieten kwamen pas daarna. Hij had ons eerst in het denkschema gebracht waarin die regels betekenis kregen.

De verklaring nam niets weg van de grilligheid van de lijn of van de schoonheid van het inzicht. Zij maakte zichtbaar hoe exactheid uit een proces van steeds fijnere benadering kon ontstaan. Dat voorbeeld vormt voor mij het tegendeel van het terrified-effect. Een verschijnsel wordt niet noodzakelijk minder waardevol wanneer wij begrijpen door welke formele structuur het mogelijk wordt. Verklaring kan onttoveren wanneer zij een onhoudbare mythe wegneemt, maar zij kan tegelijk nieuwe verwondering oproepen over de samenhang die zij blootlegt.

9 De premisse van menselijke uitzonderlijkheid

De overgang van verbazing naar existentiële angst berust vaak op een impliciete normatieve premisse. Menselijke waardigheid zou afhankelijk zijn van eigenschappen die geen ander wezen of systeem bezit. Zolang alleen mensen kunnen redeneren, spreken, componeren en betekenisvolle verbanden leggen, lijken die vermogens een uitzonderlijke plaats te rechtvaardigen. Wanneer een machine vergelijkbare prestaties levert, verdwijnt dan niet alleen een technische grens. Ook het argument voor menselijke superioriteit verliest zijn exclusieve grondslag.

Deze redenering verwacht exclusiviteit met waarde. Uit het feit dat een vermogen niet uitsluitend menselijk is, volgt niet dat het bij mensen waardeloos wordt. Vogels konden vliegen voordat mensen vliegtuigen bouwden en rekenmachines rekenen sneller dan mensen zonder dat wiskundig inzicht daardoor betekenisloos is geworden. Wanneer een programma muziek produceert die op Chopin lijkt, bewijst dat hoogstens dat bepaalde waarneembare kenmerken van Chopins werk technisch kunnen worden benaderd. Het bewijst niet dat zijn leven, ervaring of scheppingsproces eenvoudig was.¹³

Een resultaat kan langs verschillende wegen tot stand komen. De waarneembare overeenkomst tussen twee producten bewijst daarom geen identiteit van de onderliggende processen. Tegelijk kan het verschil in ontstaanswijze niet worden gebruikt om de waarneembare prestatie eenvoudig weg te definiëren. Als een systeem een nieuwe analogie correct toepast, kan men on-

¹³ Een vergelijkbare ervaring had ik op zestienjarige leeftijd bij het lezen van Simon Vestdijks *De glanzende kiemcel*. Beschouwingen over poëzie (1950), gebaseerd op lezingen die hij in 1943 in Sint-Michielsgestel hield. In het eerste hoofdstuk bespreekt Vestdijk Keats' beroemde regel uit *Endymion*: "A thing of beauty is a joy for ever." Hij vergelijkt die met de variant "A thing of beauty is a constant joy." Op het eerste gezicht lijkt het verschil in poëtische schoonheid zich moeilijk rationeel te laten verklaren. Vestdijk laat echter zien dat daarvoor ten minste zeven concrete redenen zijn aan te wijzen, onder meer in betekenis, herkomst woorden, woordkeuze, klank, ritme, zinsbouw en rijm mogelijkheden. Zijn punt is principieel: wat wij aanvankelijk ervaren als een ongrijpbaar of zelfs mysterieus esthetisch verschil, kan bij nadere analyse voor een belangrijk deel blijken te berusten op identificeerbare technische kenmerken. De grens tussen het onuitsprekelijke en het verklaarbare verschuift zodra de analyse verfijnder wordt. Het gevoel wordt onttoverd. Zie Simon Vestdijk, *De glanzende kiemcel* ('s-Graveland: De Driehoek, 1950), hoofdstuk I.

derzoeken welke kennis en generalisatie daarvoor nodig waren. Het antwoord kan verschillen van de menselijke werkwijze, maar dat verschil maakt de prestatie niet ongedaan.

AI maakt vooral zichtbaar dat mensen minder goed uit een product kunnen afleiden welk innerlijk proces het heeft voortgebracht dan zij geneigd waren te denken. Een ontroerend muziekstuk lijkt een ontroerde maker te veronderstellen. Een vloeiend antwoord lijkt een spreker te veronderstellen die weet wat hij bedoelt. Wanneer een kunstmatig systeem hetzelfde effect bereikt, ontstaat de neiging te kiezen tussen twee overhaaste conclusies: het systeem moet een menselijke innerlijke wereld bezitten, of het resultaat kan geen werkelijke betekenis hebben. Beide conclusies gaan verder dan de waargenomen prestatie rechtvaardigt.

Een steviger humanisme hoeft menselijke waardigheid niet te funderen op een monopolie op cognitieve prestaties. Mensen hebben belangen, relaties, kwetsbaarheid en een geschiedenis. Zij maken deel uit van sociale en politieke verbanden waarin rechten en verantwoordelijkheden worden toegekend. Die normatieve positie verdwijnt niet wanneer een machine een tekst schrijft of een diagnose beter voorspelt. Wie waardigheid laat afhangen van technische onnavolgbaarheid, maakt haar juist kwetsbaar voor iedere volgende uitvinding.

10 Betekenis zonder begrip

Maar zelfs wanneer men Mitchells onderscheid tussen menselijke en kunstmatige intelligentie volledig aanvaardt, blijft een andere vraag open: waarom zou het voor de betekenis en bruikbaarheid van een geproduceerde tekst noodzakelijk zijn dat het systeem zelf begrijpt wat het zegt? Moet het bovendien zelfbewust zijn voordat zijn output met die van mensen kan worden vergeleken?

Stel dat wij ons nog in het tijdperk vóór de huidige kunstmatige intelligentie bevinden. Mensen voeren gesprekken, schrijven rapporten, geven adviezen en proberen elkaar te overtuigen. Wij beoordelen zulke uitingen in de eerste plaats op hun inhoud. Is het argument geldig? Klopt de informatie? Is het advies bruikbaar? Is de redenering consistent? Begrijpen wij wat ermee wordt bedoeld? Aan de gesproken woorden of geschreven tekst zelf kunnen wij echter niet waarnemen dat zij afkomstig zijn uit een bewust brein dat begrijpt wat het zegt. Wij nemen dat bij onze medemens als vanzelfsprekend aan, maar die aanname maakt geen onderdeel uit van de tekst zelf.

Dat is geen onbelangrijk onderscheid. Een geschreven advies verschijnt voor de lezer als een zelfstandig object. De tekst wordt input voor diens eigen cognitieve verwerking. De lezer interpreteert de woorden, toetst de argumenten, verbindt ze met bestaande kennis en kent er betekenis aan toe. De mentale toestand van de schrijver is daarbij niet rechtstreeks toegankelijk. Wij kunnen aannemen dat hij begreep wat hij schreef, maar voor de beoordeling van de inhoud beschikken wij uiteindelijk alleen over wat er staat.

Bij kunstmatige intelligentie lijkt plotseling een andere eis te worden gesteld. Wanneer een systeem een begrijpelijk en samenhangend antwoord produceert, wordt gevraagd of het systeem zélf de betekenis van zijn woorden begrijpt. Als dat niet het geval is, zou er iets wezenlijks ontbreken. Maar zelfs indien men zonder voorbehoud aanneemt dat een AI-systeem geen bewustzijn bezit en niets begrijpt in de menselijke betekenis van dat woord, volgt daaruit niet dat de geproduceerde tekst daarom voor een menselijke ontvanger betekenisloos zou zijn.

De betekenis die een tekst voor een lezer heeft, is immers niet identiek aan de mentale toestand van degene die de tekst heeft voortgebracht. Een argument kan geldig zijn zonder dat de-

gene die het formuleerde de geldigheid volledig doorzag. Een beschrijving kan correct zijn zonder dat de producent ervan een innerlijke voorstelling bezit van hetgeen wordt beschreven. Een berekening kan juist zijn zonder dat het mechanisme dat haar uitvoert begrijpt wat een getal is. De afwezigheid van begrip bij de producent bewijst daarom niet de afwezigheid van betekenis in de output.

Dat betekent niet dat bewustzijn, begrip en intentionaliteit irrelevant zijn. Zij kunnen juist van groot belang zijn wanneer wij niet alleen de inhoud van een afzonderlijke uiting beoordelen, maar ook de bron ervan. Wij willen bijvoorbeeld weten waarom iemand iets beweert, op welke kennis een oordeel berust, of iemand verantwoordelijkheid kan dragen voor een advies, of hij bewust misleidt, ironisch spreekt, belangen heeft, of onder veranderende omstandigheden tot een ander oordeel zou komen. Ook betrouwbaarheid over langere tijd, consistentie van overtuigingen en morele of juridische verantwoordelijkheid veronderstellen eigenschappen die niet eenvoudig uit één geïsoleerde tekst kunnen worden afgeleid.

Daarom moeten twee vragen zorgvuldig uit elkaar worden gehouden. De eerste luidt of een systeem zelf begrijpt wat het zegt, bewustzijn bezit of bedoelingen heeft. De tweede luidt of de output die het produceert begrijpelijk, correct, betekenisvol en bruikbaar kan zijn voor een menselijke ontvanger. De eerste vraag betreft de aard van het systeem; de tweede de eigenschappen van het voortgebrachte resultaat en de verwerking daarvan door de gebruiker. Uit een negatief antwoord op de eerste vraag volgt niet automatisch een negatief antwoord op de tweede.

Juist op dit punt dreigt in veel discussies over kunstmatige intelligentie een verborgen verschuiving plaats te vinden. Eigenschappen die bij mensen relevant zijn voor de beoordeling van de spreker of schrijver worden gebruikt als criterium voor de beoordeling van de tekst zelf. Daardoor krijgt menselijke subjectiviteit een rol die zij bij gewone communicatie zelden expliciet hoeft te vervullen. Wij vragen doorgaans niet of een goed advies goed is omdat de adviseur zich bewust was van ieder betekenisniveau in zijn formulering. Wij beoordelen het advies op zijn inhoud en gevolgen. Pas wanneer wij de adviseur als bron, actor of verantwoordelijke persoon willen beoordelen, wordt diens innerlijke positie relevant.

Voor kunstmatige intelligentie zou hetzelfde onderscheid moeten gelden. Het ontbreken van bewustzijn of begrip kan belangrijke consequenties hebben voor verantwoordelijkheid, betrouwbaarheid, intentie en controle. Maar het zegt op zichzelf nog niets beslissends over de inhoudelijke kwaliteit van een afzonderlijk antwoord. Hetzelfde geldt voor de schoonheid van een compositie of een gedicht. Wie die twee niveaus door elkaar haalt, maakt van menselijke innerlijkheid een voorwaarde voor betekenis waar dat logisch niet noodzakelijk of relevant is.

Maar door die herkomst vervolgens wél relevant te maken, gebeurt er nog iets anders. Wanneer van een spreker, schrijver, analist of adviseur wordt verlangd dat hij expliciet vermeldt dat een tekst geheel of gedeeltelijk met behulp van AI tot stand is gekomen, krijgt de herkomst van de tekst invloed op de beoordeling van de inhoud. Niet omdat de argumenten, feiten of formuleringen daardoor veranderen, maar omdat aan de productie ervan een label wordt toegevoegd. Het werkt enigszins als het bordje ‘eerste kwaliteit’ bij de appels van de groenteman: nog voordat wij zelf hebben geproefd of beoordeeld, wordt onze waardering gestuurd door informatie over de herkomst of classificatie van het product.

Voor zulke transparantie kunnen soms goede redenen bestaan, maar wanneer de vermelding van AI-gebruik wordt verlangd zonder dat die herkomst voor de inhoudelijke beoordeling rele-

vant is, kan zij ook een andere functie krijgen. Het label markeert dan bij voorbaat een onderscheid tussen menselijke en kunstmatige productie en nodigt uit om dezelfde tekst anders te waarderen zodra bekend is hoe zij tot stand kwam. In die zin kan de nadruk op herkomst worden opgevat als een subtiële manier om de gedachte van menselijke uitzonderlijkheid te beschermen.

11 Wat Hofstadters angst kan rechtvaardigen

Het zou te gemakkelijk zijn Hofstadters reactie uitsluitend toe te schrijven aan gekrenkte ijdelheid van de menselijke soort. Zijn angst bevat een ernstiger mogelijkheid. Kunst is voor hem geen los product, maar een vorm van menselijk contact. In een compositie ontmoeten wij niet alleen een patroon van noten. Wij menen er aandacht, ervaring, keuze en een persoonlijk gezichtspunt in te horen. Wanneer een programma een overtuigend gelijkend product voortbrengt, kan het lijken alsof die relatie tussen maker en luisteraar op een misverstand berustte.

Die zorg verdient respect. Generatieve systemen kunnen de culturele omgeving vullen met werk waarvan herkomst en bedoeling onduidelijk zijn. Zij kunnen stijl losmaken van auteurschap, imitatie goedkoper maken dan langdurige ontwikkeling en het vertrouwen aantasten dat een tekst, beeld of muziekstuk het spoor van een concrete menselijke poging bevat. Dit zijn geen louter filosofische ongemakken. Zij raken aan erkenning, inkomen, eigendom, verantwoordelijkheid en de instituties waarmee een samenleving culturele waarde toekent.

Toch rechtvaardigen deze gevolgen niet de conclusie dat menselijke creativiteit door de technische ontdekking zelf is ontmaskerd als iets onbeduidends. Zij stellen vragen over herkomst, selectie en waardering. Een handgeschreven brief kan betekenis houden wanneer een machine dezelfde woorden zou kunnen formuleren, omdat haar betekenis mede ligt in wie haar schreef en waarom. Zo kan kunst waarde ontleen aan een menselijke geschiedenis zonder dat ieder waarneembaar kenmerk uitsluitend door mensen kan worden voortgebracht. Hofstadters angst wordt daarmee begrijpelijker, maar zij verandert van een metafysische schrik in een cultureel en institutioneel probleem.

12 Darwins merkwaardige omkering

Daniel Dennett stelde dat Darwins evolutietheorie voor veel mensen een gevaarlijk idee is, omdat zij een diep verankerd verklaringsschema omkeerde. Ontwerp leek een ontwerper te vereisen. Een oog, een vleugel of een complex organisme moest volgens die intuïtie het product zijn van een vooruitziende intelligentie. Darwin liet zien hoe variatie, erfelijkheid en selectie functionele organisatie kunnen voortbrengen zonder dat het eindresultaat vooraf is bedacht. De doelmatigheid is werkelijk, maar de bedoeling hoeft niet aan het proces vooraf te gaan.¹⁴

Dennett beschreef die omkering later als *competence without comprehension*. Een proces kan doelmatig functioneren zonder te begrijpen wat het doet. De natuurlijke selectie kent geen toekomstig ontwerp, maar brengt organismen voort die buitengewoon goed aan hun omgeving zijn aangepast. Begrip hoeft daarom niet aan iedere vorm van competentie vooraf te gaan. In de menselijke evolutie kan begrip zelf zijn voortgekomen uit een lange geschiedenis van competenties die aanvankelijk zonder begrip functioneerden.¹⁵

¹⁴ Daniel C. Dennett, *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (New York: Simon & Schuster, 1995).

¹⁵ Daniel C. Dennett, *From Bacteria to Bach and Back: The Evolution of Minds* (New York: W. W. Norton, 2017). De formulering "competence without comprehension" wordt hier gebruikt als analytische verbinding met de dis-

Kunstmatige intelligentie herhaalt deze omkering op een terrein dat directer aan ons zelfbeeld raakt. Wij zijn gewend een goede tekst te verklaren uit het begrip van de schrijver en een juiste interpretatie uit de wereldkennis van degene die interpreteert. Grote taalmodellen tonen dat indrukwekkende competentie ook kan ontstaan uit training op enorme hoeveelheden menselijk materiaal, zonder dat vooraf vaststaat dat het systeem op menselijke wijze begrijpt. De volgorde van verklaring wordt daardoor onzeker. Misschien hoeft een deel van de competentie niet uit reeds aanwezig begrip te worden afgeleid.

Daarmee is niet bewezen dat een taalmodel niets begrijpt. Evenmin is bewezen dat het bewustzijn bezit. De voorzichtige conclusie luidt dat menselijk begrip geen aangetoonde noodzakelijke voorwaarde is voor iedere prestatie die wij eerder alleen uit menselijk begrip meenden te kunnen verklaren. Dat is het gevaarlijke idee. Het tast niet de waarde van een bewezen natuurwet of een kunstwerk aan, maar wel een vertrouwd schema waarin competentie altijd van bovenaf door een begrijpende actor aan het proces wordt opgelegd.

13 Mitchell na ChatGPT

Het nieuwe voorwoord bij Mitchells boek maakt deze discussie scherper. Zij probeert haar eerdere oordeel niet te beschermen door de recente vooruitgang te bagatelliseren. Zij noemt de ontwikkeling enorm en zelfs verbluffend. Voorbeelden die zij in de eerste editie gebruikte om het gebrek aan gezond verstand en visueel begrip te demonstreren, worden door ChatGPT en vergelijkbare systemen nu aanzienlijk beter opgelost.¹⁶

Mitchell beschrijft hoe zij ChatGPT vragen stelde over een verhaal dat in haar boek als test voor alledaagse kennis diende. Het systeem beantwoordde de vragen correct en legde zijn antwoorden overtuigend uit. Omdat het verhaal mogelijk in de trainingsgegevens had gestaan, probeerde zij ook nieuwe verhalen. Ook daar bleken de antwoorden sterk. Bij foto's kon het systeem niet alleen objecten benoemen, maar ook een waarschijnlijke voorgeschiedenis en emotionele ondertoon herkennen die eerdere beeldsystemen misten.

Zij schrijft daarom dat deze systemen tekst en beelden op een diepe, mensachtige manier lijken te begrijpen. Dat is een belangrijke concessie omdat zij niet slechts een hogere score op een benchmark vaststelt. Zij erkent gedrag dat raakt aan het centrale criterium van haar boek: kennis combineren, analogieën gebruiken en nieuwe situaties interpreteren. Tegelijk handhaaft zij haar oordeel dat de fundamentele beperkingen nog niet zijn overwonnen. Haar antwoord luidt not yet.

Mitchell wijst op onmenselijke fouten, beïnvloedbaarheid door de formulering van vragen, verzonden details en kwetsbaarheid voor doelbewust misleidende invoer. Deze tekortkomingen zijn reëel. Zij beperken de betrouwbaarheid en maken controle noodzakelijk, vooral wanneer een fout grote gevolgen heeft. Ook menselijke adviezen kunnen echter controle vereisen. De filosofische vraag is daarom wat de onvolkomenheden van AI bewijzen. Uit feilbaarheid volgt niet zonder meer afwezigheid van begrip. Mensen herinneren zich gebeurtenissen verkeerd, laten zich door suggestieve vragen beïnvloeden en vullen ontbrekende informatie soms met grote zekerheid in.

cussie over AI.

¹⁶ Mitchell, Artificial Intelligence (2025), nieuw voorwoord, over taal- en beeldinterpretatie door ChatGPT en beperkingen van grote taalmodellen.

Het relevante verschil ligt daarom niet in de simpele aanwezigheid van fouten, maar in hun patroon, frequentie, contextgevoeligheid en herstelbaarheid. Een foutpatroon kan aantonen dat een model anders functioneert dan een mens. Het kan ook laten zien dat een systeem voor een bepaalde toepassing onbetrouwbaar is. Zonder een aanvullende theorie volgt er echter niet uit dat ieder begrip ontbreekt. Anders dreigt begrijpen een grensbegrip te worden dat telkens wordt verplaatst zodra een eerder beslissend geachte prestatie technisch beschikbaar komt.

Mitchell signaleert zelf dat de betekenis van AGI in de loop van de tijd smaller is geworden. In oudere voorspellingen ging het om machines die konden lezen, grappen maken, een auto smeren en zich fysiek en sociaal in de wereld redden. Tegenwoordig wordt AGI vaak beperkt tot cognitieve of intellectuele taken. Daarachter ligt de omstreden veronderstelling dat zuivere intelligentie kan worden losgemaakt van lichamelijke ervaring en sociale deelname. De vraag of dat kan, moet worden onderzocht. Zij mag niet stilzwijgend als een neutrale meetlat worden gebruikt.

14 Begrip als geleidelijk vermogen

Een deel van het debat loopt vast doordat begrijpen als een ondeelbare eigenschap wordt behandeld. Een systeem begrijpt werkelijk of het begrijpt helemaal niets. In de dagelijkse beoordeling van mensen gebruiken wij het begrip minder absoluut. Iemand kan een formule correct toepassen zonder haar afleiding te doorzien. Een ander begrijpt het wiskundige bewijs, maar niet de fysische betekenis. Een kind kan een sociale situatie goed aanvoelen zonder te kunnen uitleggen waarom.

Voor AI is een vergelijkbare ontleding productiever. Kan het systeem relevante verschillen herkennen, kennis naar een nieuwe situatie overbrengen, een redenering corrigeren, onzekerheid aangeven en de gevolgen van een handeling voorspellen? Kan het uitleggen welke informatie voor een conclusie beslissend was en verandert het zijn antwoord op een passende manier wanneer de omstandigheden wijzigen? Deze vragen leveren toetsbare antwoorden op. Begrip kan daarna worden gebruikt als samenvattende aanduiding voor een vastgesteld patroon van vermogens, maar niet als een onzichtbare substantie die eerst moet worden aangenomen voordat enige prestatie serieus mag worden genomen.

Deze benadering voorkomt zowel onderschatting als overschatting. Een taalmodel kan in een gesprek een indrukwekkend begrip tonen en elders bezwijken onder een kleine wijziging van de opgave. Het juiste oordeel is dan niet dat het in het eerste geval menselijk bewustzijn bezat en in het tweede geval slechts een papegaai bleek. Het oordeel moet specifiekere zijn: welke generalisatie lukte, welke mislukte, hoe stabiel is de prestatie en onder welke omstandigheden is het gedrag betrouwbaar genoeg voor het beoogde gebruik?

Dat functionele onderzoek beantwoordt niet automatisch de vraag naar subjectieve ervaring. Een systeem kan een taak begrijpen in de betekenis dat het flexibel en contextgevoelig handelt, terwijl onzeker blijft of er iets is wat het voor dat systeem betekent om die taak uit te voeren. Het bewustzijnsvraagstuk blijft daarmee open. Voor veel praktische toepassingen is het echter niet noodzakelijk eerst te beslissen of een machine een innerlijk leven heeft. Een brugberekening, medische voorspelling of tekstuele analyse moet vooral worden beoordeeld op geldigheid, betrouwbaarheid en passend gebruik.

15 Praktisch risico en metafysische angst

De kritiek op het terrified-effect mag niet worden verward met achteloosheid over AI. Een ondoorzichtig systeem kan discriminatie reproduceren, overtuigend onjuiste informatie verspreiden of beslissingen beïnvloeden zonder dat iemand verantwoordelijkheid neemt. Bedrijven en staten kunnen de technologie gebruiken om macht te concentreren. Mensen kunnen afhankelijk worden van systemen waarvan zij de beperkingen niet kennen. Naarmate systemen zelfstandiger handelen, neemt bovendien het belang toe van controleerbare doelen, begrensde bevoegdheden en de mogelijkheid om ingrijpen te stoppen of terug te draaien.

Deze gevaren worden niet helderder wanneer zij worden vermengd met angst voor het verlies van menselijke uitzonderlijkheid. De praktische vraag luidt wie een systeem inzet, met welk doel, op basis van welke gegevens, onder welk toezicht en met welke verantwoordelijkheid voor fouten. De metafysische vraag luidt of een machine werkelijk denkt, begrijpt of voelt zoals wij. Die tweede vraag kan filosofisch van groot belang zijn, maar zij is niet voor iedere toepassing de beslissende veiligheidsvraag.

Een systeem hoeft geen mens te zijn om schade te veroorzaken. Het hoeft evenmin een mens te zijn om nuttig te worden. De eis dat eerst menselijke innerlijkheid moet zijn aangetoond voordat een prestatie telt, is daarom zowel te streng als te zwak. Zij is te streng omdat een niet-menselijk proces reële kennis en bruikbare resultaten kan opleveren. Zij is te zwak omdat een systeem dat in een gesprek zeer menselijk overkomt nog steeds onbetrouwbaar, manipuleerbaar of gevaarlijk kan worden ingezet.

Ook voor regulering is dit onderscheid van belang. Wetgeving die concrete schade wil voorkomen, moet aangrijpen bij bevoegdheden, toepassingen, aansprakelijkheid, controle en aantoonbare risico's. Zij moet niet stilzwijgend proberen een symbolische grens rond menselijke intelligentie te beschermen. Een samenleving kan besluiten bepaalde menselijke rollen niet volledig te automatiseren, bijvoorbeeld omdat verantwoordelijkheid of persoonlijk contact daar essentieel wordt gevonden. Dat is een normatieve en institutionele keuze. Zij hoeft niet te worden vermomd als de empirische bewering dat een machine principieel nooit iets kan begrijpen.

16 Van vervanging naar vergroting

Aan het einde van haar nieuwe voorwoord stelt Mitchell haar sterkste vraag: wat willen mensen eigenlijk van intelligente machines? Willen wij systemen die mensen zoveel mogelijk vervangen, of systemen die menselijke vermogens vergroten en taken uitvoeren die mensen niet kunnen of niet willen doen? Met die vraag verplaatst zij het debat van een voorspelling naar een keuze. De technische mogelijkheid bepaalt niet zelfstandig welke maatschappelijke toepassing wenselijk is.¹⁷

De inrichting van arbeid, onderwijs, zorg en bestuur bepaalt of AI vooral wordt gebruikt om kosten te verlagen en beslissingen te centraliseren, of om mensen meer inzicht en handelingsmogelijkheden te geven. Zelfs wanneer een machine bepaalde cognitieve taken beter uitvoert, volgt daar niet uit dat een organisatie de mens uit het proces behoort te verwijderen. Omgekeerd is het evenmin vanzelfsprekend dat een taak door mensen moet worden behouden wanneer een systeem haar aantoonbaar veiliger en nauwkeuriger uitvoert. De beslissing vereist een

¹⁷ Mitchell, Artificial Intelligence (2025), afsluiting van het nieuwe voorwoord, over het vervangen dan wel vergroten van menselijke vermogens.

beoordeling van doelen, risico's, verantwoordelijkheden en de menselijke waarden die in de betreffende praktijk werkelijk op het spel staan.

De bruikbaarste maatstaf is daarom niet of een systeem een volmaakt kunstmatig mens is. Wij moeten vragen wat het betrouwbaar kan, voor welk doel, onder welke voorwaarden en met welke mogelijkheden tot controle en correctie. Wanneer menselijk inzicht, verantwoordelijkheid of contact noodzakelijk is, moet het systeem daaraan dienstbaar blijven. Wanneer een formele methode mensen bevrijdt van tijdrovend werk of hun waarneming vergroot, is het geen verlies dat die methode anders functioneert dan een mens.

Misschien ligt hier ook een meer fundamentele praktische grens. Mensen handelen altijd binnen relaties met andere mensen en kunnen, ook zonder die bedoeling te hebben, schade veroorzaken of de persoonlijke integriteit van anderen aantasten. Een maatschappelijke rechtsorde verlangt daarom dat verantwoordelijkheid voor zulke gevolgen kan worden toegewezen. Het gebruik van een intelligent systeem mag niet leiden tot een situatie waarin niemand meer aanspreekbaar is omdat een beslissing 'door de AI' zou zijn genomen. Er moet steeds een mens, organisatie of andere rechtspersoon zijn die verantwoordelijkheid draagt voor de inzet van het systeem, de voorwaarden waaronder het wordt gebruikt en, waar redelijkerwijs mogelijk, de gevolgen daarvan.

Daarmee verschuift ook hier de relevante vraag. Niet of het systeem zelf verantwoordelijkheid kan dragen, maar wie verantwoordelijk is voor het gebruik ervan. Voor de inhoudelijke beoordeling van een advies kan het onbelangrijk zijn of de kennis door menselijke analyse, statistische berekening of kunstmatige intelligentie tot stand kwam. Voor het rechtens iemand verantwoordelijk stellen kan de wijze waarop dat advies is verkregen relevant zijn, bijvoorbeeld wanneer iemand een systeem zonder voldoende controle gebruikte of bekende beperkingen negeerde. Kunstmatige intelligentie hoeft dus geen zelfstandig moreel of juridisch subject te zijn om maatschappelijk bruikbaar te kunnen zijn; noodzakelijk is vooral dat haar inzet geen ontsnappingsroute uit menselijke verantwoordelijkheid wordt.

17 Conclusie

Hofstadters angst vormt een intrigerend filosofisch probleem omdat zij botst met het inzicht waarvoor zijn eigen werk zo invloedrijk is geworden. GEB liet zien hoe betekenisvolle patronen, zelfverwijzing en mogelijk zelfs een ik kunnen ontstaan uit de organisatie van onderdelen die deze eigenschappen afzonderlijk niet bezitten. Wanneer kunstmatige systemen onverwachte cognitieve vermogens ontwikkelen, ligt daarin geen eenvoudige bevestiging van alle verwachtingen over AI. Het is wel een verschijnsel waarvoor Hofstadters denken ons beter had voorbereid dan vrijwel ieder ander.

De terrified-reactie wordt begrijpelijk wanneer menselijke intelligentie wordt beschouwd als de grondslag van menselijke superioriteit. AI bedreigt dan niet alleen werk, veiligheid of bestaande instituties, maar ook een zelfbeeld. De dreiging ontstaat doordat eigenschappen die als exclusief golden mogelijk reproduceerbaar blijken. Die reproduceerbaarheid bewijst niet dat menselijke intelligentie eenvoudig is en evenmin dat machines op dezelfde wijze denken of voelen. Zij toont wel dat het waarneembare resultaat minder rechtstreeks naar een unieke menselijke innerlijkheid verwijst dan vaak werd aangenomen.

Daarmee is AI een gevaarlijk idee in de betekenis die Dennett aan Darwin gaf. Zij keert een vertrouwd verklaringsschema om. Competentie hoeft misschien niet altijd te beginnen met

menselijk begrip en een complex resultaat hoeft niet vooraf door een begrijpende ontwerper te zijn uitgedacht. De passende reactie op die mogelijkheid is onderzoek, begripsvorming en nauwkeurige beoordeling. Angst wordt rationeel wanneer zij een concreet gevaar identificeert. Zij wordt misleidend wanneer zij een bedreigd privilege van de mens voor een technisch veiligheidsprobleem houdt.

De machine heeft de mens niet kleiner gemaakt. Zij heeft slechts een ondeugdelijk argument voor menselijke grootheid weggenomen.

Wij hoeven intelligentie niet mysterieus te houden om haar bijzonder te vinden en kunst niet onverklaarbaar te verklaren om erdoor ontroerd te worden. Menselijke waardigheid hoeft niet te worden beschermd door iedere verrassende machineprestatie buiten de grens van werkelijk begrip te plaatsen. De ontdekking dat een formeel of statistisch systeem meer kan dan wij verwachtten, kan dezelfde soort verbijstering oproepen als de oneindig klein wordende rechthoeken van mijn wiskundeleraar. Zij nodigt uit tot een nieuw denkschema.

Mijn bezwaar tegen Hofstadters angst doet daarom niets af aan mijn bewondering voor hem. GEB heeft mijn denken fundamenteel veranderd en blijft een van de oorspronkelijkste onderzoeken naar de mogelijkheid van een mechanistische geest. Juist daarom wil ik Hofstadters eigen intellectuele gereedschap op zijn reactie toepassen. De verschijning van onverwachte kunstmatige competentie hoeft hem niet filosofisch te overvallen. Zij maakt zichtbaar hoe vruchtbaar en hoe gevaarlijk voor mensen blijkaar zijn eigen idee altijd al was.

Literatuur

- Bergman, Frank. 2026. "Top Investor Sounds Alarm: AI Elites Are Trying to Create a 'God-Like' Intelligence to Rule Humanity." Slay News, 7 juni. <https://slaynews.com/top-investor-sounds-alarm-ai-elites-trying-create-god-like-intelligence-rule-humanity/>.
- Birhane, Abeba, en Marek McGann. 2024. "Large Models of What? Mistaking Engineering Achievements for Human Linguistic Agency." *Language Sciences* 106: 101672. Openbare auteursversie: <https://doi.org/10.48550/arXiv.2407.08790>.
- Brynolfsson, Erik, Ajay Agrawal, Anton Korinek en Tom Cunningham, organisatoren. 2026. "We Must Act Now: A Statement on AI's Transformation of the Economy." Juli. <https://www.wemustactnow.ai/>.
- Cope, David. 1989. "Experiments in Musical Intelligence (EMI): Non-linear Linguistic-based Composition." *Interface* 18 (1–2): 117–139. <https://doi.org/10.1080/09298218908570541>.
- Dennett, Daniel C. 1995. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster.
- Dennett, Daniel C. 2017. *From Bacteria to Bach and Back: The Evolution of Minds*. New York: W. W. Norton.
- Forrester, Jay W. 1969. *Urban Dynamics*. Cambridge, MA: MIT Press.
- Genewein, Tim, e.a. 2026. "From AGI to ASI." arXiv:2606.12683. <https://doi.org/10.48550/arXiv.2606.12683>.
- Google en Ipsos. 2026. Google/Ipsos Multi-Country AI Survey 2026. 15 januari. <https://www.ipsos.com/en-us/google-ipsos-multi-country-ai-survey-2026>.
- Grace, Katja, e.a. 2025. "Thousands of AI Authors on the Future of AI." *Journal of Artificial Intelligence Research* 84. <https://doi.org/10.1613/JAIR.1.19087>.
- Hofstadter, Douglas R. 1979. *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books.
- Mitchell, Melanie. 2019. *Artificial Intelligence: A Guide for Thinking Humans*. New York: Farrar, Straus and Giroux.
- Mitchell, Melanie. 2025. *Artificial Intelligence: A Guide for Thinking Humans* (With a New Preface). New York: Picador.
- OpenAI. 2026. "From Asking to Doing: How the World Is Putting ChatGPT to Work." 6 augustus. <https://openai.com/index/how-the-world-is-putting-chatgpt-to-work/>.
- Roosen, Christopher. 2023. "It Doesn't Matter If Artificial Intelligence Can Write a Novel—It's the Wrong Thing to Automate." 2 augustus. <https://www.christopherroosen.com/blog/2023/8/2/story-engine-word-machines-writing-automation>.
- Stanford Institute for Human-Centered Artificial Intelligence. 2026. The 2026 AI Index Report. Hoofdstuk "Economy". <https://hai.stanford.edu/ai-index/2026-ai-index-report/economy>.
- Vestdijk, Simon. 1950. *De glanzende kiemcel: Beschouwingen over poëzie*. 's-Graveland: De Driehoek.

