Niet onze AI

Gabriella Obispa & Laurens Vreekamp

Voorbij de beloften van Big Tech

VANDUUREN MEDIA

NIET ONZE AI

Voorbij de beloften van Big Tech

Gabriella Obispa
Laurens Vreekamp

VANDUUREN
MEDIA

ISBN: 978-94-6356-430-4
NUR: 801
Trefw.: AI, management
Redactie en zetwerk: Van Duuren Media, Culemborg/Barcelona
Illustraties: Dirma Janse
Layout: Yannick Mortier
Fotografie: Nicola Diamondaras
Druk: Wilco, Amersfoort

Eerste druk: november 2025

Dit boek is gedrukt op een papiersoort die niet met chloorhoudende chemicaliën is gebleekt. Hierdoor is de productie van dit boek minder belastend voor het milieu.

'Een gesprek over technologie kan niet beginnen met de technologie zelf.'

— Ruha Benjamin, socioloog en professor Afro-Amerikaanse Studies aan de Universiteit van Princeton.

Inhoud

Voorwoord vii

Deel 1: I – Hun AI **1**

1 AI en de mens 5

2 Wat bezielt de eindbazen? 33

3 Consequenties van 'hun' AI 53

Deel 1: II – Onze AI **67**

4 Geen natuurverschijnsel 71

5 Positie en representatie 93

6 Comfort, principes en reflectie 125

7 Meer keuzes, andere smaken 145

Deel 1: III – Jouw AI **163**

8 Oefenen met iets anders 167

9 Beginnetjes van beweging 193

10 Durf te dromen 213

Verantwoording AI-gebruik bij schrijven 227

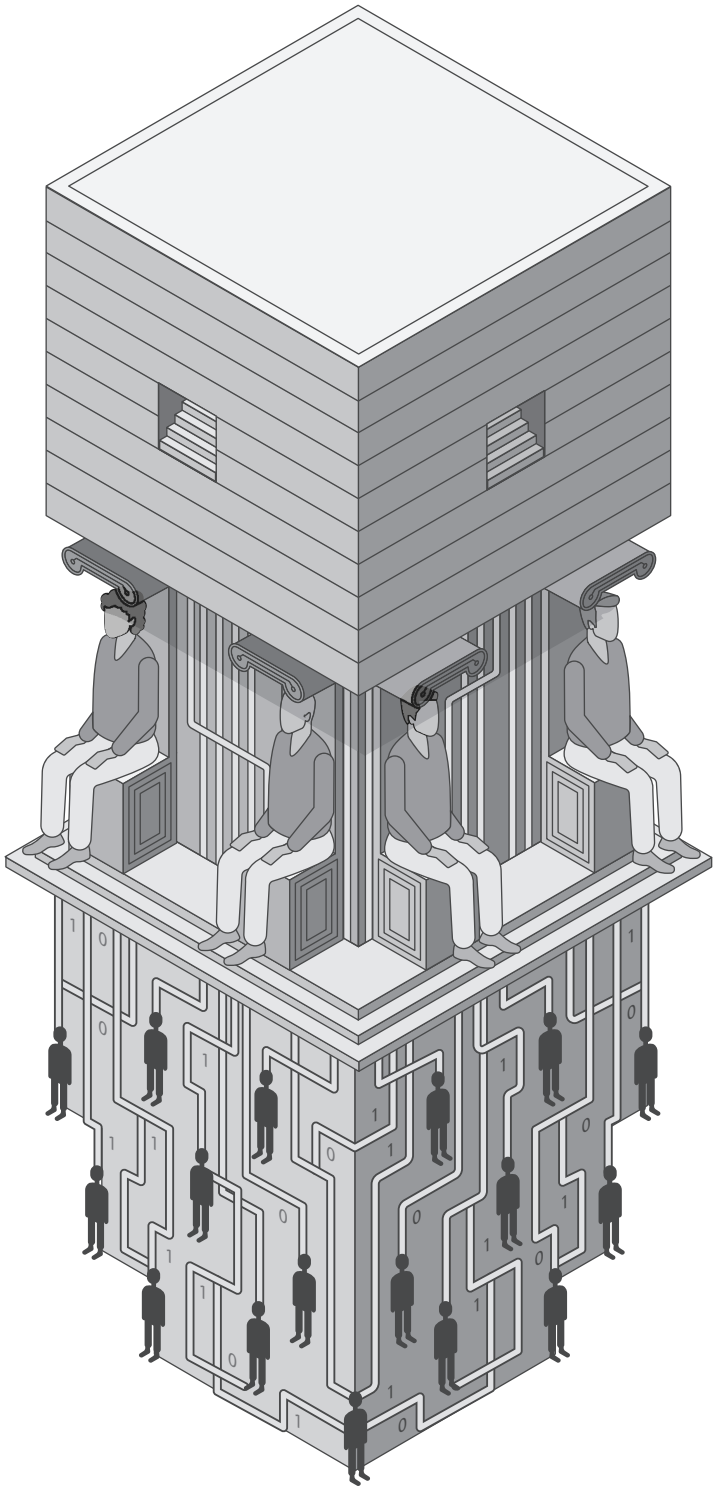
Dankwoord 228

Bijlage: 25 vragen om te stellen alvorens AI toe te staan in je organisatie 231

Flowchart voor doordachte AI-inzet 236

Bronnen 239

Index 241



I – Hun AI

Wellicht heb je Alexander Klöpping op televisie zien vertellen hoe gebruikers persoonlijke relaties aangaan met ChatGPT. Ze beschouwen de chatbot als metgezel, therapeut, psycholoog of liefdespartner. Niet gek, want AI-systemen presenteren zich bewust als intelligente wezens. Ze spreken je aan met 'ik' en doen geen moeite om te verhullen wat ze in feite zijn: computersoftware gebaseerd op statistiek en wiskunde, getraind met grote hoeveelheden materiaal van het internet. In ronkende marketingteksten van AI-aanbieders wordt gesproken over 'begrijpen', 'denken' en 'redeneren'. Het is in het belang van de aanbieders om je dat te laten geloven, zodat jij je hart en ziel bij ze uitstort.

We zouden ondertussen toch geleerd moeten hebben dat wij, gebruikers, de techbedrijven betalen met onze data? Door inzage te geven in ons online gedrag en door het opgeven van allerlei persoonlijke informatie voorzien we Big Tech al bijna drie decennia van gegevens die voor hen miljarden waard zijn. 'Als je er niet voor betaalt, ben je zelf het product' is hier de bekende wijsheid. Daar wordt met AI een nieuwe dimensie aan toegevoegd. Wanneer we onze vragen en diepste geheimen delen met de chatbots, krijgen we er weliswaar soms bruikbare antwoorden voor terug. Maar de bouwers winnen er onevenredig veel meer mee dan wij. Omdat de belangen zo groot zijn, heeft Big Tech een verhaal gesponnen dat breed gedragen wordt, maar misleidend is. Want de onderliggende software kan namelijk helemaal niet denken, niet begrijpen en niet redeneren, en is ook niet in staat om tot een vermeende superintelligentie te komen. Hoe scheef de belangen zijn, en hoe diep we in het verhaal gezogen zijn, laten we in dit deel zien. We helpen je de mythes van de werkelijkheid te onderscheiden.

“De mens is volgens Silicon Valley veelal een statische, homogene gebruiker, geïsoleerd van de omgeving en van anderen”

Judith Zoë Blijden

Oprichter Is This Art Maybe
en Senior Adviseur Hooghiemstra
& Partners

1

AI en de mens

In het voorwoord schreven we dat we het niet te veel willen hebben over wát AI is. Toch vinden we het wel handig om *iets* over de werking ervan uit te leggen. Er zijn grofweg twee soorten AI op dit moment. De nieuwste en populairste versie noemen we generatieve AI. Als mensen het nu over AI hebben, bedoelen ze vaak deze variant. Het is het soort AI waar chatbots gebruik van maken: ze genereren (eigenlijk: reproduceren) tekst, code en pixels op basis van onderliggende taalmodellen (daar komen we zo op). De 'oude' versie van AI ken je misschien als 'machine learning'. Het is de vorm die goed is in classificeren, clusteren, samenvatten en gesproken teksten transcriberen. Deze variant genereert zelf geen nieuwe inhoud, maar analyseert vooral bestaande data. Het verwerkt inhoud en kan daarmee voorspellingen doen. Deze versie maakt dus vooral onderscheid tussen het een en het ander, en wordt daardoor 'discriminatieve AI' genoemd.

Een andere belangrijke term is 'algoritme'. Mensen gebruiken de termen AI en algoritmen vaak door elkaar. Het zijn twee verschillende dingen, die beide werken met regels. Het grote verschil tussen AI en algoritmen is dat bij die laatste de regels worden opgesteld door mensen. Bij AI-systemen stelt de machine de regels op, op basis van door mensen gegeven voorbeelden. Algoritmen bestaan uit voorgeprogrammeerde instructies die de computer precies vertellen wat het moet doen. 'Als dit zo is, dan doe je dat'. Wat voor data deze regelgebaseerde algoritmen ook krijgen, ze blijven hetzelfde functioneren, zonder zich zelfstandig aan te passen. Ze worden niet als 'intelligent' beschouwd, maar volgen simpelweg hun voorgeprogrammeerde regels. [1]

AI-systemen doen voorspellingen die gebaseerd zijn op wiskunde en statistiek. De zogeheten LLMs, ofwel grote taalmodellen (*large language models*), liggen ten grondslag aan chatbots als ChatGPT. De modellen werken met kansberekeningen op basis van analyse van veelvoorkomende teksten die in de trainingsdataset zitten. Dat gaat om miljarden documenten en teksten – enorm grote aantallen die alleen de krachtigste computers aankunnen. De modellen zetten tijdens hun training woorden en delen van zinnen om in getallen. Elk woord(deel) krijgt een nume-

rieke waarde en ook de relaties tussen woorden en zinsdelen worden vastgelegd in getallen. 'Winter', 'koud' en 'muts' hebben bijvoorbeeld onderling een sterke relatie met elkaar, omdat ze vaak bij elkaar staan in de teksten in de trainingsdata. 'Steen, 'stekker en 'stupide' zullen minder vaak bij elkaar voorkomen, dus de kans dat je ze alle drie in een enkele zin nodig hebt, is heel klein. Alles wat de modellen analyseren tijdens hun training, wordt in cijfers uitgedrukt. Dit omdat machines met cijfers heel goed, snel en op grote schaal kunnen rekenen. Als alle relaties zijn berekend – 'gewogen' in jargon – worden ze vastgelegd in het uiteindelijke model. Dit is het wiskundige deel. (In essentie werkt het proces bij beeldgeneratoren hetzelfde, alleen worden dan in plaats van woorden de kleurenpixels in getallen omgezet en worden de pixelrelaties vastgelegd.)

Als je een vraag stelt aan een chatbot, worden de woorden en tekens die je gebruikt in die vraag (de 'prompt') ook omgezet naar getallen. Die getallen worden vergeleken met de getallen in het model. Als je een antwoord van ChatGPT krijgt, is die tekst het resultaat van kansberekeningen: er is door het model voorspeld welke woorden en zinsdelen logisch passen en aansluiten bij jouw prompt. Dit is het statistiekdeel: dit en dat komt vaak bij elkaar voor, dus de kans is groot dat jij dit ook bedoelt en dit bij het antwoord past. Taalmodellen rekenen en voorspellen dus eigenlijk met taal. Ze kennen geen zekerheid, hebben geen kennis over hun inhoud en zijn ook geen zoekmachines. Taalmodellen reproduceren grammaticaal correcte zinnen, maar hebben geen notie van betekenis of feiten in het resultaat. Wat ze doen is waarschijnlijkheid bieden op basis van een voorspelling. Het antwoord van een chatbot toont dus het resultaat van tientallen kansberekeningen en voorspelt dat de teruggegeven woorden, zinnen, pixels en regels code bij elkaar horen, omdat ze zo vaak in de trainingsdata dicht bij elkaar stonden.

Professor in de taalkunde Emily Bender vergelijkt taalmodellen met papier-maché, waarbij de maker papier (vaak kranten) gebruikt vanwege de fysieke eigenschappen ervan. De krantenteksten die zodoende in het kunstwerk terechtkomen zijn bijzaak. Zo werken LLMs niet met de betekenis van de tekst die ze verwerken en uitvoeren, maar alleen met de vorm – letters, woorden en zinnen. "LLMs zijn niets meer dan modellen voor de verdeling van de woordvormen in hun trainingsgegevens." [2] De Amerikaanse techniekfilosoof Shannon Vallor omschrijft LLMs als

"in de basis voorspellende wiskundige modellen die uitingen van menselijke intelligentie imiteren, zoals redeneren." Chatbots hebben volgens haar de vervelende gewoonte antwoorden te verzinnen die "statistisch plausibel, maar feitelijk onjuist" zijn. Dit is een logisch, rechtstreeks gevolg van de techniek waarmee de modellen worden ontwikkeld, zo zul je hierna ontdekken.

Ingrediënten voor intelligentie

In het geval van therapie en liefde is het onmogelijk om ons menselijk wezen te vatten in software. Dit heeft twee fundamentele oorzaken: we hebben de neiging om de mogelijkheden van software te *over*schatten en de complexiteit van het menszijn te *onder*schatten. Sterker nog: die complexiteit is zo groot dat verschillende, samenwerkende wetenschapsdisciplines deze nog altijd niet volledig weten te doorgronden, laat staan beschrijven. Onze psychologische levens, onze karakters en onze interpersoonlijke en sociale interacties zijn simpelweg niet te vatten en om te zetten in voor computers werkbare input: datapunten. Toch is dat wat techbedrijven ons doen geloven en beloven: de systemen zouden volgens hen een aantal dingen kunnen die levende wezens ook doen: 'denken' en 'logisch redeneren'. Het is goed om nog even naar de basis te gaan en de volgende vraag te stellen: hoe wordt kunstmatige intelligentie ontwikkeld?

De meeste en bekendste generatieve-AI-systemen van dit moment zijn de chatbots die werken met de taalmodellen. We roepen ze aan via populaire programma's als ChatGPT, Google Gemini, Microsoft Copilot of Anthropics Claude. In Europa biedt het bedrijf Mistral LeChat aan en vanuit China maakte de wereld begin 2025 kennis met DeepSeek. Om AI-software zoals taalmodellen te kunnen bouwen, heb je veel (dure) ingrediënten nodig. Zie het als een recept met een bijbehorend boodschappenlijstje. Slechts een handjevol bedrijven zoals OpenAI, Google, Microsoft en Meta heeft toegang tot alle benodigde ingrediënten om AI-modellen te bouwen: de financiering, de benodigde kennis om AI-software te bouwen, menselijk talent, de toegang tot immense rekenkracht in datacenters, het bezit van kostbare hardwareonderdelen (zoals de zeer gewilde grafische processoren van Nvidia) en *last but not least*: de talloze terabytes aan gegevens, afkomstig van sociale

media, het world wide web, uit boeken en online bibliotheken. Al die data vormen de grondstof om de machines mee te trainen, zodat er na training een taalmodel als resultaat is. Je chatbot roept dat taalmodel aan om een antwoord te genereren.

AI-systemen zijn erop getraind, zoals gezegd, om patronen te vinden in die grote berg gegevens en die vervolgens vast te leggen in statistische modellen. En omdat er elke dag een enorme hoeveelheid gegevens nieuw beschikbaar komt, moeten deze modellen zichzelf continu hertrainen; ze kunnen immers niet de toekomst voorspellen. Bovendien is er een wedstrijd gaande tussen de bedrijven achter de verschillende modellen. Ze willen allemaal alsmaar nieuwe updates uitbrengen met taalmodellen en chatbots die nog weer 'slimmer' zijn. Wie uiteindelijk de bot bouwt die door de meeste mensen wordt geadopteerd is spekkoper. Maar het bouwen van iedere nieuwe versie van een taalmodel kost met gemak enkele miljoenen dollars. Tel daar ook de hoeveelheid energie bij op die nodig is voor de opslag van datagegevens, die keer op keer wordt verbruikt om modellen te ontwikkelen en de energie die het kost om onze gesprekken met de chatbots te voeren.

Dan hebben we het nog niet eens over andere, indirecte kosten van AI, soms eufemistisch aangeduid als 'verborgen'. Denk aan de winning van de benodigde hoeveelheid aardse materialen, zoals kobalt en lithium. Al dat materiaal is nodig voor het vervaardigen van hardware en het bouwen en operationeel houden van de datacenters. De delving van al die aardse materialen heeft impact op ons klimaat. Tegelijkertijd zijn er ook kosten die eigenlijk wel gemaakt zouden moeten worden, maar nu niet worden vergoed door de techbedrijven. Ze betalen namelijk niet voor de rechten van de gegevens die ze van het internet schrapen. De grote AI-bedrijven schenden de naburige en auteursrechten van muzikanten, schrijvers, ontwerpers, journalisten, filmmakers, fotografen en illustratoren. Ze gebruiken zonder toestemming, zonder naamsvermelding en zonder compensatie de materialen van anderen om hun AI-modellen mee te trainen. In Silicon Valley waarschuwen insiders dat als rechters oordelen dat hiervoor betaald moet worden, een groot deel van onze AI-toepassingen niet meer zou kunnen bestaan. [3]

Andere kosten gaan over de offers die datawerkers brengen. Dat zijn de mensen die AI-trainingsdata voorzien van labels en die de antwoorden van chatbots beoordelen. Vaak zijn het mensen in lagelonenlanden die vierkantjes trekken om objecten in foto's, om bijvoorbeeld een stropdas, een overhemd en pantalon te leren herkennen. Ze voorzien teksten en videofragmenten van labels (spannend, humoristisch, gewelddadig) en bekijken real-time videobeelden van zelfrijdende auto's om in te grijpen waar nodig. Ze beoordelen of chatbotantwoorden goed passen bij de prompts van gebruikers en geven aan of socialemediaberichten kinderporno bevatten of tot geweld oproepen. Het zijn deze mensen die uren achtereen, vaak slecht betaald het ergste van het web moeten filteren, zodat het ons nooit bereikt via AI-gegenereerde resultaten. Zonder hen zou er geen bruikbaar materiaal voor AI zijn om mee getraind te worden. Journalist Jeroen van Bergeijk, die undercover ging als datawerker, hoorde dat journalisten, tekstschrijvers en redacteuren de laatste tijd actief benaderd worden door recruiters om zogeheten AI-trainer te worden, juist vanwege hun algemene kennis en goede beheersing van de Nederlandse taal.

Datawerkers krijgen veelal slecht betaald bij het doen van dit geestdodende en mentaal uitputtende werk. Het is daarbij ook zeer traumatiserend door de bedenkelijke, onsmakelijke, gruwelijke en/of strafbare inhoud die de datawerkers daarbij tegenkomen. En het is onzeker en stressvol werk omdat het ad hoc komt en gaat en zo onvoorspelbare inkomsten oplevert. Om hoeveel mensen het precies gaat is lastig te achterhalen, maar een schatting van de Wereldbank is dat honderden miljoenen mensen zich bezighouden met het trainen van AI-systemen of het labelen van data. In Europa gaat het volgens onderzoek van Claartje ter Hoeven, hoogleraar communicatie- en informatiewetenschappen aan de Universiteit Utrecht, om miljoenen mensen die voor deze dataplatforms werken.

Perfecte uitgangspunten

Al deze benodigde ingrediënten om AI-systemen te ontwikkelen zijn gekoppeld aan specifieke lagen of stappen in een mondiale keten van hard- en softwarebouwers en -leveranciers. Ze vormen samen iets dat tegenwoordig de *tech stack* heet. Haroon Sheikh, lid van de Wetenschappelijke Raad voor de Regering, legt het zo uit: "In de wereld van de tech-

nologie kijkt men vaak naar objecten volgens het model van een stapel, in het Engels een stack. Een telefoon bouw je bijvoorbeeld op door verschillende lagen op elkaar te stapelen, die allemaal met elkaar samenwerken: de laag van de chips, de besturingssoftwarelaag waar de programma's op kunnen draaien, nog een laag van de zogeheten interface, die ervoor zorgt dat je als gebruiker de telefoon kan bedienen. Dan de laag van de apps, waarmee je kan zoeken, betalen, appen, twitteren en daarboven die van de netwerken, zodat het apparaat kan communiceren met het internet." [4]

Wereldwijd zijn er diverse spelers op de markt die verschillende lagen van de stack bedienen. Nederland heeft met ASML bijvoorbeeld een goede positie op de computerchiplaag. Maar het zijn vooral Amerikaanse Big Tech-bedrijven die de gehele stack verder volledig beheersen. Zij hadden en hebben een perfecte positie door al zo'n dertig jaar geleden de eerste internetgebaseerde hard- en software te ontwikkelen en te vermarkten. Denk hierbij aan Googles toepassingen voor zoeken, de smartphones en tablets van Apple en Meta's socialemediaplatforms zoals Facebook en Instagram. Inkomsten genereren ze grotendeels met de verkoop van advertenties en de verkoop van onze data. Andere inkomstenbronnen vormen de abonnementsgelden om de digitale diensten advertentievrij te gebruiken. Ondertussen zijn vooral ook clouddiensten een grote inkomstenbron van Big Tech-bedrijven. Met de cloud bedoelen we de opslag van digitaal materiaal op computerservers in datacenters, waar we als gebruikers betalen om onze bestanden veilig op te slaan en er overal en altijd snel bij te kunnen. Het is ook in de cloud waar al die AI-modellen worden getraind en draaien. Apple laat zichzelf erop voorstaan dat onze privacy en de veiligheid van onze gegevens hun hoogste prioriteit hebben. Met het creëren van zogeheten digitale ecosystemen willen techbedrijven ons zo veel binnen hun eigen omgeving houden. Want dankzij het ontginnen, verfijnen, herverpakken en verkopen van ons digitale gedrag binnen hun eigen systemen, alles gevat in data, zijn ze enorm rijk en groot geworden. Die data vertellen veel over wie wij zijn en registreren wat wij doen online.

Kijk eens naar de jaarcijfers en beurskoersen van Apple, Amazon, Alphabet (moederbedrijf van Google), Microsoft of Meta uit de laatste boekjaren. Bedenk eens hoe ze met hun wereldwijde bereik en het aantal

toepassingen dat ze ontwikkelen en onderhouden, die 24/7 in de lucht zijn, miljarden gebruikers over de hele wereld bedienen. Let eens op de overnames die ze doen en deden. Kijk naar het aantal medewerkers dat de techbedrijven gezamenlijk in dienst hebben. Na zo'n dertig jaar internet, smartphone en sociale media is het voor niemand meer een verrassing dat datahandel enorm veel geld oplevert. Degene die informatie over gebruikers en hun gedrag slim weet te koppelen aan adverteerders, kan veel geld vragen voor die uniek gecreëerde markt(plaats)positie.

Wie die informatie vervolgens omzet naar slimme en doelgerichte campagnes, kan slimmer en meer producten en boodschappen verkopen dan de concurrent. Maar dat niet alleen. Met analytische data over gebruikers en hun gedrag, kun je de publieke opinie beïnvloeden (Brexit), verkiezingen winnen (Obama, Trump), mensen desinformeren (corona), hypes creëren en overheidsbeleid bepalen. Herinner je je nog dat AI-systemen leren op basis van patronen in grote hoeveelheden data? Dan realiseer je je dat de data die techbedrijven in handen hebben, naast om te verkopen, ook waardevol is in het licht van AI. Want als al die data in grote aantallen aanwezig zijn op je eigen servers, dan heb je de perfecte grondstof om AI-software mee te trainen. Dan blijk je wel heel erg goed gepositioneerd om AI-software te gaan ontwikkelen.

In dat licht moet je ook kijken naar wat WeTransfer, een bestandsuitwisselingsdienst, recentelijk deed. Het wijzigde de voorwaarden om daarmee toestemming te verkrijgen om al het uitgewisselde materiaal tussen gebruikers te mogen inzetten voor AI-modeltraining. Zo ook moet je het verzoek zien van Meta, dat we als gebruikers in mei 2025 kregen op Facebook en Instagram. We dienden volgens het bedrijf expliciet aan te geven als we *niet* wilden dat onze foto's en berichten gebruikt werden als AI-trainingsmateriaal. Dat Meta dit verzoek deed en ons er actief op wees, leek voorbeeldig gedrag. Volgens privacykenners zou dit juist omgekeerd moeten zijn: AI-bedrijven zouden óns toestemming moeten vragen om ons materiaal te gebruiken. De huidige Europese wet- en regelgeving voorziet hier (nog) niet in. De Nederlandse Autoriteit Persoonsgegevens (AP) zag zich dan ook genoodzaakt om gebruikers actief in alle soorten media en via socialemediaplatforms, op te roepen om vooral geen toestemming te geven aan Meta. AP-vicevoorzitter Monique Verdier waarschuwde "Het risico is dat je als gebruiker de controle over jouw persoonsgegevens verliest. Kort-

om, wat gebeurt er precies met jouw gegevens en waarom heeft Meta nu juist die gegevens nodig? Je hebt ooit iets op Instagram of Facebook gepost en die gegevens zitten straks in dat AI-model. Zonder dat je precies weet wat daarmee gebeurt." [5]

Op 27 mei 2025 startte Meta met het trainen van hun nieuwe AI-model. Met jouw gegevens eenmaal in dat AI-model, zijn ze er niet zo maar weer uit te krijgen. Hoewel de AP gebruikers opriep zelf actie te ondernemen, benadrukte ze dat de verantwoordelijkheid voor naleving van privacywetgeving primair bij Meta ligt.

De marketing van magie

Technologiebedrijven doen er alles aan om hun producten en diensten zo goed mogelijk te verkopen; het zijn ware marketingmachines. Ze kleden hun beloften plezierig aan en de gepromote mogelijkheden van AI-software verbloemen bepaalde motieven en aspecten. Wat bijvoorbeeld buiten beschouwing blijft, is hoe tastbaar onderdelen van de technologie zijn. Het is een mythe dat AI-systemen alleen in de 'cloud' bestaan; ze verbruiken natuurlijke hulpbronnen. Denk aan de lagen van de tech stack met siliciumchips, kilometers aan kabels, en gebouwen vol datacenters die veel stroom en koeling nodig hebben. AI-software heeft heel veel mensen nodig om überhaupt te functioneren. Kate Crawford, hoofdonderzoeker bij Microsoft Research en auteur van het boek *The Atlas of AI*, merkte jaren geleden al op dat AI helemaal niet intelligent is, en ook niet kunstmatig, maar vooral mensenwerk is: "En dan ook nog eens grotendeels weinig uitdagend werk. Als onderdeel van het trainen van AI-systemen zijn grote hoeveelheden laaggeschoolde arbeid nodig", [6] zei ze tegen het Britse nieuwsblad *The Guardian* over het eerdergenoemde datawerk. Het zijn allemaal zaken die volgens Crawford angstvallig "achter een 'gordijn worden weggehouden' door de AI-bedrijven." Dit omdat het immers het idee van de magie van hun systemen ernstig tenietdoet.

De marketing voor AI-toepassingen en chatbots komt goed naar voren in visuele uitingen en woordgebruik van Big Tech. Zo is het meestgebruikte grafische symbool om AI-functies binnen apps en software weer te geven, het vonkpictogram. "Het geeft een gevoel van magie", zo

komt uit onderzoek naar voren van de Universiteit van Straatsburg naar taal- en beeldgebruik door AI-bedrijven. "De grafische elementen zoals de vonk worden daarbij vaak aangevuld met superlatieven en termen rondom kracht en magie – 'Krachtige assistent' (bij Elon Musks chatbot Grok), 'Ontdek de kracht' (Adobe Photoshop) en 'Duik samen met ons in prestaties' (Apple). Qua kleurgebruik wordt paars veelvuldig ingezet." Kleurspecialist Marion Lamarque licht toe: "In de beeldende kunst staat paars voor magie, maar het wordt ook geassocieerd met innovatie. [Het] voegt een gevoel van zachtheid toe en geeft de magie van AI een geruststellend gezicht, vooral in combinatie met een hemelsblauw kleurverloop, dat rustgevend is en consensus uitstraalt." [7]

Naast de eerder genoemde ingrediënten die AI-ontwikkelaars nodig hebben, blijken ook voor het verspreiden van de marketingboodschap de techbedrijven wederom perfect gepositioneerd: het communicatief talent en de pr-kracht is al in huis en ze bezitten zelf de platforms, kanalen en apps waarmee ze miljarden mensen eenvoudig en haast kosteloos weten over te halen om direct te ervaren wat AI kan. Daarbij laten ze je voortdurend de voordelen van het werken met hun AI-toepassingen 'ervaren'. Zo kun je met Googles Gmail al jaren op door AI-gegenereerde slimme antwoorden klikken of complete zinnen laten afschrijven. Je iPhone foto-app herkent automatisch huisdieren en personen die je eerder al van een naam hebt voorzien. Snapchat kreeg een eigen AI-chatbot en Microsoft Word helpt je met het samenvatten van je tekst.

De onderzoekers van de Universiteit van Straatsburg zijn kritisch: "Door visuele verwijzingen en associaties met magie op te roepen, positioneren bedrijven AI als een multifunctioneel hulpmiddel, zonder ooit te zeggen wat het nu wel en niet kan. Deze grafische onduidelijkheid bevordert de nadruk op AI in al onze interfaces, zonder onderscheid te maken tussen de grote verschillen in gebruik. In feite gebruiken bedrijven hetzelfde pictogram voor AI-gebaseerde functies die niets met elkaar gemeen hebben. De AI-knop van Zoom is bijvoorbeeld totaal anders dan die van Google, maar visueel lijken ze op elkaar. De grafische standaardisatie van sterk uiteenlopende functionaliteiten creëert een grijs gebied dat gunstig is voor de aanbieders van AI-software. Als we als gebruiker niet precies weten waar we het over hebben, kunnen we er

ook niet echt iets van verwachten. We worden verrast en hebben moeite om de effecten ervan op waarde te kunnen schatten."

Om je meer begrip te geven van wat er achter de schermen van AI gebeurt, leggen we een paar van die termen uit en ontkrachten we de magie.

Een belangrijke term die je vaak tegenkomt rondom AI, is die van het neurale netwerk. Dit soort netwerken zijn in opzet geïnspireerd op kennis over de werking van onze hersenen. Hier komt de gedachte dus vandaan dat deze software zou kunnen denken als wij. Maar de werking van onze hersenen is zelfs voor neurowetenschappers nog altijd grotendeels een mysterie. Wat AI-ingenieurs ontwikkelden en neurale netwerken noemden, is gebaseerd op onvolledige kennis over hersenen en neuronen. Bovendien moesten de ontwikkelaars die interpretaties omzetten in een wiskundige benadering, omdat digitale machines alleen met getallen kunnen werken.

Het toegepaste softwareresultaat van neurale netwerken is men *deep learning* gaan noemen. Dat heet zo omdat software gebaseerd op neurale netwerken op meerdere, diepe lagen is gestoeld. Elke laag kijkt naar een ander patroon in de data, en geeft een getal door als resultaat. Net zo lang totdat alle lagen allerlei patronen hebben gevolgd, met een becijferde uitkomst tot gevolg. Om je een paar voorbeelden te geven: via al die lagen wordt voorspeld dat bij 'macaroni' en 'tomaat' ook het woord 'Parmezaanse kaas' hoort, of voor vleeseters 'gehakt'. Middels deep learning weet een AI-systeem ook of een afbeelding een muffin bevat of een bagel, of dat een AI-gegenereerd Sinterklaasgedicht rijmt volgens AA-BB-schema.

Mensen makkelijk, computers gecompliceerd

Alhoewel de inzet van neurale netwerken in technologie tot veel doorbraken heeft geleid, en voorspellingen en berekeningen in AI-software beter werden sinds het inzetten van deep learning, mogen we AI niet gelijkstellen aan de mens. Nu je weet hoe AI-systemen in essentie worden gebouwd en hoe ze werken, is dat op geen enkele manier te vergelijken met hoe menselijke intelligentie zich ontwikkelt en werkt. Ons 'denken' bestaat uit veel meer dan cognitieve intelligentie. We hebben

ook emotionele intelligentie. Uitgaan van alleen die twee elementen doet ons mensen nog steeds tekort. We hebben de neiging om de complexiteit van de mens te onderschatten. Intelligentie bereik je namelijk niet met het nabootsen van ideeën over hoe onze hersenen zouden werken. Rudy van Belkom, directeur van de Stichting Toekomstbeeld der Techniek zegt: "Artificiële Intelligentie (AI) ontwikkelt zich met name op het gebied van taalverwerking. Hoe indrukwekkend tools als ChatGPT ook zijn, taalverwerking vormt slechts een klein deel van menselijke intelligentie. Dat zit niet alleen in je hoofd (belichaamde cognitie), maar wordt gevormd door onderdelen van het gehele lichaam (fysieke ervaringen, bewegingen en waarnemingen). Zo ontwikkelen kinderen ruimtelijk inzicht door met blokken te bouwen. Intelligentie ontstaat door interactie met de omgeving. Representaties in onze geest krijgen pas betekenis door de koppeling met sensomotorische ervaringen (zoals voelen, ruiken en proeven). Een kind leert het woord 'warm' niet door alleen er een definitie van te lezen, maar door iets warms aan te raken. Intelligentie is sociaal ('theory of mind'). Intelligentie hangt samen met het besef dat de eigen opvattingen, verlangens en emoties kunnen afwijken van die van een ander. Door doktertje te spelen, trainen kinderen hun inlevingsvermogen." [8]

Computerwetenschappers, wiskundigen, technisch ingenieurs en programmeurs zagen als eerste de concrete en mogelijke toepassingen en opbrengsten van AI. Kenmerkend voor dit type academicus is dat ze veelal in de oplossingsstand staan. Ze lossen problemen op door dingen te onderzoeken, experimenten uit te voeren en systemen te bouwen. Computers en software vormen de bepalende instrumenten tijdens het ontwikkelproces. De problemen die ze proberen op te lossen, tellen alleen wanneer zowel de ingenieurs als de machines ze herkennen en begrijpen – anders zouden ze het immers niet eens zien als een probleem. En, de oplossing moet in potentie hoge economische waarde hebben. Zo volgt het uitgangspunt voor ingenieurs dat een probleem alleen relevant is wanneer het met technologie opgelost kan worden. Dat is al met al een zeer beperkend uitgangspunt voor een technologie zoals AI, die de hele wereld zou gaan overnemen. In jargon wordt dit fenomeen complexiteitsreductie genoemd. Het soort oplossingen dat hieruit voortkomt omschrijven critici als technosolutionisme.

We nemen je ter illustratie even mee naar het onderwijs. Dit om duidelijk te maken hoe kwalijk het is als er naar een onderwerp of probleem wordt gekeken vanuit een beperkt blikveld. Bij onderzoek naar technologie in de klas wordt leren en 'naar school gaan' door softwareaanbieders vaak gereduceerd tot kennis- en informatieverwerving. Leren als iets dat primair cognitief is, in de hersenen plaatsvind. Het is een vernauwend perspectief op de realiteit en complexiteit van leren. Het laat de context van onderwijs buiten beschouwing: zintuiglijk leren, je fysiek in een klas bevinden, het omgaan met leeftijdgenoten met verschillende intelligenties, interesses, karakters, vermogens en achtergronden. Dit komt doordat de rijkheid en complexiteit van de onderwijsomgeving, hoe leren werkt, zich lastig laat vatten in softwarecode. Met andere woorden: alleen dat wat gereduceerd en gevangen kan worden in data, kan worden gedigitaliseerd. Daarmee wordt onderwijs en leren van veel betekenis ontdaan. Want, zoals gezegd, hebben kinderen voor kennisverwerving meer nodig dan toegang tot (persoonlijk afgestemde) informatie. Sociale context is minstens zo belangrijk, met daarbij een belangrijke rol voor de vele functies die een docent uitoefent: rolmodel, kennisautoriteit, motivator, verantwoordelijk voor een veilige leeromgeving. Ook de invloed van klas- en leeftijdgenoten op elkaar, bijvoorbeeld bij het stellen van normen en delen van waarden, laten zich lastig vangen in techniek alleen. Uit onderzoek van Designlab Universiteit Twente blijkt dat het door kinderen niet zomaar wordt geaccepteerd wanneer functies en rollen van de docent in digitale vorm worden aangeboden. "De meerderheid van de kinderen is duidelijk in hun oordeel dat besluitvorming niet (geheel) uitbesteed mag worden aan computers die in hun intieme sfeer komen", zo schrijven de onderzoekers [9]. Het zit in onze natuur om ons graag en vooral tot mensen te verhouden. In relatie tot machines hebben we daar natuurlijk gezien andere verwachtingen van en niet dezelfde relatie mee.

Tegelijkertijd hebben veel mensen het idee dat programmeerwerk en computercode iets zeer complex en ondoorgrondelijks is. In de media en in fictieverhalen worden developers vaak geportretteerd als genieën met een uniek talent om de wereld redden. Felienne Hermans, professor aan de Vrije Universiteit, ziet dat de publieke perceptie rond AI en programmeren zodanig is dat het een bovenmenselijk moeilijke taak lijkt. De film

The Matrix komt vaak terug in de beeldspraak van programmeurs en heeft ook bij het publiek postgevat. Ze schreef in de Volkskrant: "Iedereen denkt nu dat AI-algoritmen zo ingewikkeld zijn dat de programmeurs zelf ze niet eens snappen. Gewone mensen moeten zich niet met AI bemoeien, zei bijvoorbeeld Eric Schmidt, voormalig CEO van Google. En dit verhaal blijven de media herhalen. Vakgenoten spinnen er garen bij dat algoritmen moeilijk te snappen zijn, want dat is goed voor hun marktpositie. We moeten het niet te makkelijk maken, codes horen ingewikkeld te zijn. Maar in een wereld waarin programmeren, grotendeels door het vakgebied zelf, als extreem ingewikkeld en ondoordringbaar wordt voorgespiegeld, kan bijna niemand de bullshit doorzien." [10] Hermans heeft, mede door dit gegeven gedreven, zelf een eenvoudig toegankelijke programmeertaal ontwikkeld voor niet-programmeurs. Ze geeft daarnaast een dag per week les in coderen en programmeren op het Codasium, onderdeel van het Rotterdams Lyceum Kralingen.

AGI, de evolutie van AI

AI zou dus kunnen 'redeneren'. De software werkt met neurale netwerken. 'Superintelligentie' is nabij. Veel mensen zijn er ondertussen van overtuigd dat AI een soort magie is, dat systemen kunnen denken, dat chatbots logisch kunnen redeneren en dat software slimmer wordt dan mensen. Daar bijkomend wordt AI gepresenteerd als iets dat inherent goed voor ons is.

Dat heeft alles te maken met de heersende opvattingen en wereldbeelden van de dominante makers achter de technologie. Kunstmatige systemen, zo beweren sommigen, zijn ondertussen niet alleen even slim als mensen, ze kunnen straks dingen die wij niet eens meer zullen begrijpen. Er worden voor de evolutie van AI grofweg vier stadia van bewustzijn gehanteerd [11]:

1. het reactieve stadium,
2. het stadium van beperkt geheugen,
3. de 'theory of mind' (dit noemde Rudy van Belkom eerder bij intelligentie), en:
4. zelfbewustzijn.

Bij het eerste stadium, het reactieve, kun je denken aan de eerste 'smart speakers' zoals Siri, Alexa en de Google Assistant die je met je stem bedient en waarbij je als gebruiker een (gesproken) antwoord terugkrijgt. Het waren computeringenieurs die de eerste werkende AI-toepassingen bedachten en ontwikkelden, voor bijvoorbeeld het zakenreizigersprobleem: vind de meest efficiënte weg naar verschillende klanten. Andere opdrachten waar vroege AI'ers aan werkten gingen over het logisch aan elkaar knopen van losse stukken bedrijfsinformatie in kennissystemen. De vroegste commerciële consumententoepassingen van AI vond je op Amazon.com: aanbevelingen voor boeken, films en andere complementaire producten. Je kunt al dit soort vraagstukken in het reactieve stadium plaatsen.

Bij het tweede stadium, beperkt geheugen, onthouden chatbots als Microsofts Copilot en OpenAI's ChatGPT eerdere prompts en conversaties – dingen die je al typend of sprekend hebt gevraagd, en die opgeslagen liggen in je chathistorie. Op basis daarvan kun je aangepaste reacties verwachten of er gericht naar verwijzen voor het krijgen van een beter antwoord.

Bij het derde stadium wordt het interessant, dat van *theory of mind*. Dat is het vermogen van een persoon om zich een idee te vormen van het perspectief van een ander. We hebben allemaal onze eigen kijk op de wereld en onze eigen emoties en reacties. Het bepaalt hoe we reageren op een situatie. Met AI zouden we dit derde stadium bereikt hebben wanneer een zelfrijdende auto in staat is om niet alleen te toeteren naar een menselijke weggebruiker die de AI-wagen afsnijdt, maar ook de boosheid, frustratie, angst en onveiligheid ervaart die ermee gepaard gaan.

We zouden het laatste stadium, zelfbewustzijn, bereiken wanneer het computersysteem gevoelens begrijpt en weet waarom een gebeurtenis een bepaalde reactie uitlokt. Dan is het ook beter in staat om reacties van anderen echt te begrijpen. Zo gaan we er als automobilist bijvoorbeeld van uit dat iemand die in een auto naar ons toetert kwaad of gefrustreerd is, aangezien we zelf ook zouden kunnen toeteren wanneer iemand in het verkeer ons hindert. Een of meer van die emoties die een zojuist afgesneden weggebruiker ervaart zou het AI-systeem dan ook

moeten ervaren. En daar bovenop zou het AI-systeem een voorkeur moeten hebben voor boos, angst of frustratie, want een zelfbewust systeem dient over een persoonlijkheid en karakter te beschikken.

Dit verste stadium – zelfbewustzijn – bereiken lijkt vanuit meerdere perspectieven onmogelijk, bijvoorbeeld bekeken vanuit sociaalpsychologische, emotionele en filosofische niveaus. AI-critici beweren terecht dat AI de meeste fundamenten van zelfbewustzijn mist. AI heeft geen lichaam. Het heeft geen besef van tijd, geen honger, geen behoefte aan rust of de wens om zich voort te planten. Het kan geen liefde, pijn, vreugde of andere emotie voelen. Het kan woorden genereren die beweren dat het een gevoel ervaart, maar dat is niet hetzelfde. AI-software werkt met elektrische impulsen; het is geen hersenchemie. [12]

Technisch gezien is bewustzijn ontwikkelen in software fundamenteel ingewikkeld. Wetenschappers moeten eerst volledig doorgronden hoe ons bewustzijn werkt en hoe we onze keuzes aanpassen aan eerdere gebeurtenissen. Techbedrijven blijven de laatste tijd dan ook weg bij termen als bewustzijn, en richten zich in hun marketingteksten bewust op metaforen die aansluiting hebben bij de hersenenmetafoor en vooral met cognitie te maken hebben: hoe ons denken werkt. Daarom lees je nu veel over chatbots die kunnen begrijpen, logisch redeneren en diepgaand onderzoek voor je verrichten. Voor de goede orde: dat kunnen de modellen en chatbots niet – ze blijven nog altijd statistische voorspellingen doen op basis van patronen die ze hebben geleerd uit hun trainingsdata. Het zijn probabilistische modellen: dat betekent dat ze waarschijnlijkheden berekenen, kansen. En waarschijnlijkheden en kansen bieden geen zekerheid. Het zijn geen deterministische, op vaste regels gebaseerde toepassingen, zoals rekenmachines dat wel bieden. Die zullen op sommen als 9+3 of 114000/12 altijd dezelfde antwoorden geven: 12 en 9500. Iedere keer weer.

Ondanks dat het duidelijk is dat het vierde stadium nog lang niet is bereikt, en misschien ook nooit bereikt zal worden, lijken veel AI-bouwers en -experts ongehinderd door de bekende tendens om de mogelijkheden en impact van technologie te overschatten. Zij voorzien zelfs een soort vijfde stadium: die van algemene kunstmatige intelligentie, kortweg

AGI genoemd (*artificial general intelligence*).[1] Dat is een vorm van kunstmatige intelligentie die zelfstandig nieuwe, en misschien wel een eigen soort kennis creëert. Voor AGI-aanhangers is het maar een kleine (denk)stap om vanuit de huidige stand en mogelijkheden (*capabilities*) van generatieve AI te komen tot een vorm van AI die de wereld echt begrijpt en in principe alles kan. Zij denken dat als je de AI-systemen en modellen maar groot genoeg maakt, het bewustzijn in AI-systemen vanzelf zal ontstaan. De software bereikt dan het stadium van AGI. Het kernidee van AGI is dat het in de meeste taken beter is dan mensen. AGI heeft dan de potentie om heel de wereld te veranderen. Maar dan moeten we wel eerst een breed gedeelde definitie van AGI hebben. Dat blijkt lastig. Er is namelijk geen consensus over wat AGI nu precies is en niemand weet zeker of het überhaupt kan bestaan.

In de nieuwsbrief *The Algorithm* van het Amerikaanse Massachusetts Institute of Technology (MIT) lezen we dat het opvalt in gesprekken over AGI dat steeds het volgende aan bod komt: "Het bouwen van zo'n ultrakrachtig systeem kan een werkelijk onpeilbare hoeveelheid middelen vereisen – geld, chips, edele metalen, water, elektriciteit en menselijke arbeid. Maar als AGI (hoe vaag ook gedefinieerd) zo krachtig is als het klinkt, dan is het alle kosten waard." [13] In de nogal hyperige paper *Sparks of AGI* [14] uit maart 2023 stelden de onderzoekers dat taalmodel GPT-4 'vonkjes van superintelligentie', AGI dus, vertoonde. Probleem is alleen dat iedereen in de AI-wereld iets anders verstaat onder AGI. OpenAI, het bedrijf achter ChatGPT, definieert AGI als "zeer autonome systemen die beter presteren dan mensen bij het meeste economisch waardevolle werk", terwijl Elon Musk zegt dat AGI AI is die "slimmer is dan de slimste mens". Dario Amodei, CEO van Anthropic, maker van chatbot Claude, denkt dat AGI meer lijkt op een "land van genieën, maar dan gevat in een datacenter".

Het is inmiddels wel duidelijk dat als er al iets van vonkjes van AGI zouden zijn, of wanneer iemand beweert AGI te hebben bereikt, het niet duidelijk is of iedereen het daarmee eens zal zijn. Toch lijken alle grote AI-bedrijven ernaar op zoek en deden vrijwel alle topmannen van AI-

[1] In 2025 kwam ook de afkorting ASI vaker op als alternatief voor AGI. ASI staat dan voor *artificial super intelligence*. Beide termen worden door elkaar heen gebruikt.

bedrijven al uitspraken over de datum waarop we het zullen bereiken. De financiële site Sherwood News zocht overigens uit of er ook AI-experts die geen man zijn dit soort voorspellingen doen, maar konden die niet vinden. [15]

Wat verder opvalt is dat al deze AGI-overtuigingen niet daadwerkelijk ingaan op de vele hobbels op de weg waar AI-onderzoek en -implementatie mee te maken krijgt. Spreek je met mensen die het AGI-geloof aanhangen, dan zijn ze niet in staat toe te lichten hoe precies de toepassingsspecifieke AI – onze huidige AI-systemen – zal overgaan in die algemene kunstmatige intelligentie. Wat ze doen is de tijdlijn voor AGI zo ver doortrekken naar de toekomst dat bestaande hobbels er simpelweg niet meer toe lijken te doen. Met andere woorden: je wordt met een AGI-kluitje in het riet gestuurd.

Gelukkig is er juist ook in Nederland uitvoerig onderzoek gedaan naar de eventuele mogelijkheid van AGI. Onderzoek van een team met Prof. Dr. Iris van Rooij, hoogleraar in computationele cognitiewetenschappen aan de Radboud Universiteit Nijmegen, toonde aan dat het realiseren van AGI met machines die alleen leren op basis van data computationeel onhaalbaar is. [16] Simpeler gezegd: AGI is niet mogelijk. De zogenaamde 'kansen' op AGI waar zo vaak over gesproken wordt, zoals ook Alexander Klöpping deed bij de talkshow Eva van de publieke omroep op 22 mei 2025, zijn dan ook niet op feiten gebaseerd, schrijft Associate Professor M. Birna van Riemsdijk van de Universiteit Twente: "Het is een vaker gebruikte retorische truc om te stellen dat zelfs bij een geringe kans op het geschetste scenario aandacht hiervoor op zijn plaats zou zijn. Met zo'n redenering kun je alles relevant verklaren zolang je maar een kans groter dan 0 poneert dat het bewaarheid wordt." [17]

Waarom zouden we die onwaarschijnlijke staat van AGI eigenlijk nastreven, horen we je denken. En hoe waardevol is AI eigenlijk als het werkelijke verhaal lang niet zo rooskleurig is als we dachten? Big Tech blijft het verheerlijkende verhaal over AI en AGI als redder van de wereld verspreiden, omdat het er groot belang bij heeft. Het heeft er vooral baat bij om financiering te blijven krijgen, om uit de handen van regulering door de overheid te blijven en zo hun macht verder uit te breiden. Ze gebruiken de vage belofte van AGI en wat dat voor de mensheid kan

betekenen, omdat er nog grotere taalmodellen ontwikkeld moeten worden, zodat de systemen echt AGI kunnen bereiken. En daarvoor zijn nog meer datacenters en grafische processoren (GPU's) en servers nodig. En die kosten geld.

Hoe we denken en praten over AI is op dit moment vooral in lijn met hoe Big Tech het graag ziet: dat AI de *gamechanger* is sinds internet en de mobiele telefoon. De Nederlandse Wetenschappelijke Raad voor het Regeringsbeleid (WRR) sprak al in november 2021 over AI als een 'systeemtechnologie', vergelijkbaar met elektriciteit en internet. Het draagt bij aan het narratief dat AI iets belangrijks, alomvattends en onvermijdelijk is. Tegen investeerders (en soms ook hun gebruikers) zeggen de bouwers van AI steevast: natuurlijk, onze systemen werken nog niet feilloos, maar we zijn onderweg naar dat AGI. Maar daar is dus wel veel geld voor nodig, want de modellen moeten doorontwikkeld kunnen worden. En om grote sommen geld binnen te halen, heb je durfinvesteerders nodig. En daarvoor heb je als techbedrijf een hype nodig.

Ex-Googler en huidig directeur van berichtenapp Signal, Meredith Whittaker, legde recent uit hoe dat werkt: "Het bedrijfsmodel van durfkapitaal moet worden gezien als een hype. Durfkapitaal kijkt naar waarderingen en groei, niet noodzakelijk naar winst of omzet. Je hoeft dus niet echt te investeren in technologie die werkt, of die zelfs winst maakt, je moet gewoon een verhaal hebben dat overtuigend genoeg is om die waarderingen te laten stijgen. Deze repetitieve en uitputtende hypecyclus is dus een kenmerk van deze sector. [...] Het is niet simpelweg zo dat een stuk technologie overhyped is; hype is een noodzakelijk ingrediënt van het huidige (zakelijke) tech-ecosysteem. We moeten onderzoeken hoe vaak de financiële prikkel voor een hype wordt beloond zonder enig echt sociaal rendement, zonder enige betekenisvolle vooruitgang in technologie, zonder dat deze tools en diensten en werelden zich ooit echt manifesteren. Dat is de sleutel tot het begrijpen van de groeiende kloof tussen het verhaal van techno-optimisten en de realiteit van onze met technologie belaste wereld." [18]