
HUMaiN
Een zoektocht naar wat AI
is en wat het met ons doet
Salem Samhoud & Freek Grootenboer

HUMaiN

Denken over AI

Salem Samhoud & Freek Grootenboer
Project HUMaiN - Een initiatief van &samhoud

INHOUD

Inleiding	3
Grondleggers	10
De drie Godfathers	11
De drie Godfathers van AI Deel 1	15
De drie Godfathers van AI Deel 2	21
De drie Godfathers van AI Deel 3	29
De Bouwers	36
De tien bedrijven	37
Wat de bouwers van AI denken	39
Van Gelijkgestemd tot Rivalen	41
De onderliggende belangen van Open Source	47
De pragmatische weg vooruit	53
Wie betaalt de transitie?	59
De Waarschuwing	66
De kracht die we voelden	67
An Inconvenient Truth 2.0	71
De Verborgen Gevaren	78
Dichterbij dan we dachten	79
De stille prijs van efficiëntie	81
De Verhouding	88
De vraag die met ons meebeweegt	89
Hoe verhouden wij ons tot AI?	93
Waar we staan en wat we gaan doen	108

Inleiding

Een gedachtereis in twee stemmen

SALEM — *Intellectueel initiatiefnemer van HUMaiN*

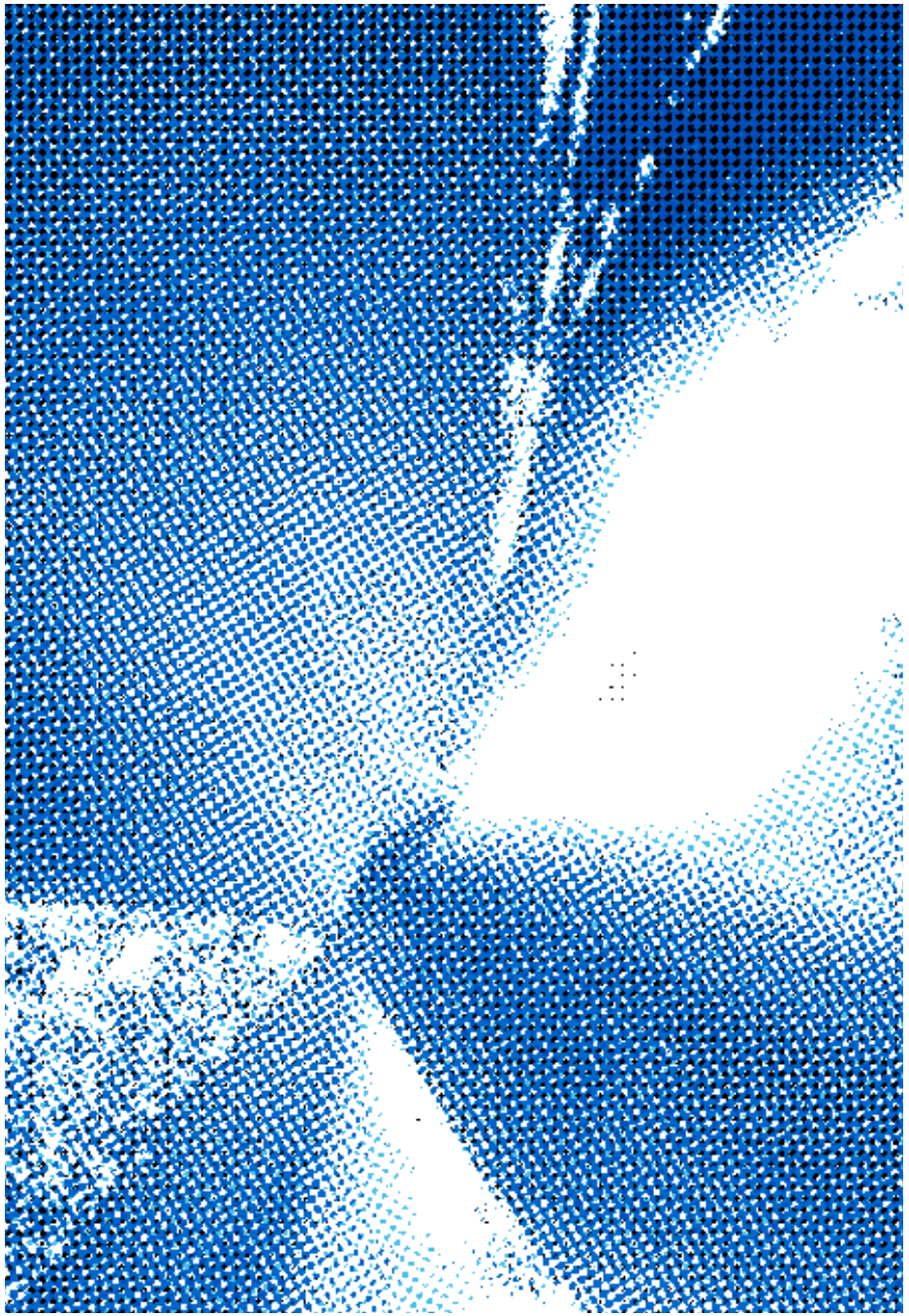
Het is zo'n dertig jaar geleden. Onze dochter was nog niet geboren. Ik was met mijn vrouw Marjan en onze zonen op vakantie in Australië. We sliepen in een tent, midden in de wildernis. Marjan en de kinderen sliepen al. Boven mij stond de hemel zo vol sterren dat het bijna onwerkelijk was.

Je voelt je op zo'n moment heel klein. Nietig, in een oneindig universum. Miljoenen sterren die fonkelden in de nacht. Ik weet nog dat ik dacht: hier moet toch meer zijn? Het kan niet zo zijn dat wij de enigen zijn. Ander leven. Bewuste wezens. Wezens die verder zijn dan wij, of gewoon anders. En meteen kwamen de vragen. Zijn ze ons goedgezind? Kunnen we ze vertrouwen? Hoe zou zo'n ontmoeting eruitzien? Ik hoopte dat ik dit ooit in mijn leven zou meemaken. Die wens is nooit meer weggegaan. Niet als angst. Als fascinatie.Mijn eerste echte ervaring met kunstmatige intelligentie kwam jaren later. Ik stond in Silicon Valley, midden in het geweld van Google en Facebook. In de marge van dat bezoek sprak ik een man bij een klein augmented reality bedrijf. De interface om met AI te praten bestaat nog niet, zei hij. Zodra die er is, verandert alles. Hij beweerde dat augmented reality de interface voor AI zou worden. AI bestond toen al, maar vrijwel niemand was er warm voor. Over dat ene woord — alles — kreeg hij volkomen gelijk. Want toen ChatGPT verscheen, was het de eerste technologie die ik zelf kon gebruiken. Niet via een jongere collega. Gewoon typen. Een vraag stellen zoals je die aan een mens stelt. En een antwoord krijgen dat klopte. Ik was onder de indruk. Gegrepen was ik nog niet. Dat kwam met een experiment. Een collega liet me op een middag honderd LinkedIn-berichten zien. Allemaal gegenereerd op basis van mijn eigen teksten.

Ze waren raak. Herkenbaar. Mijn stem, mijn gedachten, mijn ritme. Ik plaatste er één. En ik voelde me een leugenaar. Daarna ging ik verder. Ik liet de machine een analyse van mezelf maken. Een beeld van mijn persoonlijkheid en mijn manier van leidinggeven. Wat eruit kwam was verbluffend precies. Het raakte dingen die ik zelf nauwelijks onder woorden kon brengen. En het was niet alleen vriendelijk. Er stond ook scherpe kritiek in. Rake kritiek. Het bijzondere was dit. Die kritiek kon ik aannemen. Makkelijker dan wanneer een mens hem had uitgesproken. Want de machine had geen belang bij mijn antwoord. Geen ego. Geen status te verdedigen. Ze hoefde mij niet te ontzien en wilde ook niets van mij. Juist daardoor kwam de boodschap binnen. Dat moment zette een keten van vragen in gang. Die stopte niet meer.

Wat is eigenaarschap, als een machine jouw stem kan nadoen? Wat is echtheid, als het product niet te onderscheiden is van het origineel? En dieper nog: wat blijft er over van de mens, als de machine zijn meest waardevolle bijdragen kan overnemen? Ik ben geen pessimist. Naïef ben ik evenmin. Ik zag hoe de grondleggers van AI bijna allemaal begonnen met dezelfde belofte: de wereld beter maken. En ik zag hoe die belofte onder druk van geld en macht begon te schuiven. Ik zag hoe snel het ging. En ik zag hoe weinig mensen er serieus over nadachten. Zeker in het bedrijfsleven.
Terwijl de omwenteling enorm zal zijn. Banen die verdwijnen, of waarvan de inhoud radicaal verandert. Macht die verschuift. Ongelijkheid die nog groter wordt dan ze nu al is.
Er zijn vandaag invloedrijke ondernemers en denkers uit de Verenigde Staten die openlijk pleiten voor een basisinkomen en voor herverdeling. Voor belasting op digitale arbeid. Dat zulke ideeën nu in die kringen serieus klinken, zegt iets. Het wijst op een diepe zorg.

Ik besloot dat ik het wilde begrijpen. Echt begrijpen. En ik wist precies wie ik daarvoor nodig had.



FREEK — *Filosoof en schrijver, verbonden aan project HUMaiN*

Salem belde mij aan het begin van het jaar. We hadden, zoals elk jaar, het World Economic Forum op de voet gevolgd. AI stond daar al een jaar of drie hoog op de agenda. Maar deze editie was anders. Waar het eerst gold als een krachtige, opkomende technologie, was het nu bepalend voor alles. Iedereen leek hetzelfde inzicht te delen. Wie de kracht van AI weet te ontsluiten en echt weet te gebruiken, zal winnen. Dat geldt voor organisaties, voor landen, voor hele continenten. Zakelijk. Politiek. En als versterker van oplossingen voor de grote crises van onze tijd. Na het forum hadden we al besloten dat we dieper de techniek in moesten. Begrijpen wat het is. Begrijpen wat je ermee kunt.

Maar toen Salem me een paar weken later belde, ging het om meer. Hij vroeg me mee te doen aan HUMaiN. Om samen deze reis te maken. Ik dacht eerst dat we een mooi artikel zouden schrijven. Het bleek veel groter. We zouden echt de tijd nemen. Tijd en middelen om grondig na te denken. Om met specialisten te spreken. Om te lezen. Wat me overviel was zijn passie. De intense nieuwsgierigheid. En de grondigheid waarmee hij de stof in wilde duiken. Voor mij was AI tot dan vooral iets dat me intellectueel boeide. Een verschijnsel met grote invloed, verbonden met macht. Voor een filosoof eindeloos interessant. De kans om me daar volledig in te verdiepen was haast te mooi om waar te zijn.

We begonnen met studeren, lezen en interviews. We discussieerden urenlang. In het begin was het te groot. Een vloed aan vragen. En de vraag achter die vragen: stelden we de juiste? Begrepen we eigenlijk wel waar we het over hadden?
Al snel merkte ik dat ik gepassioneerd raakte. De voorspellingen over waar de techniek heen beweegt, zijn voor een groot deel onzeker. Maar als zelfs een fractie ervan uitkomt, is de impact enorm.

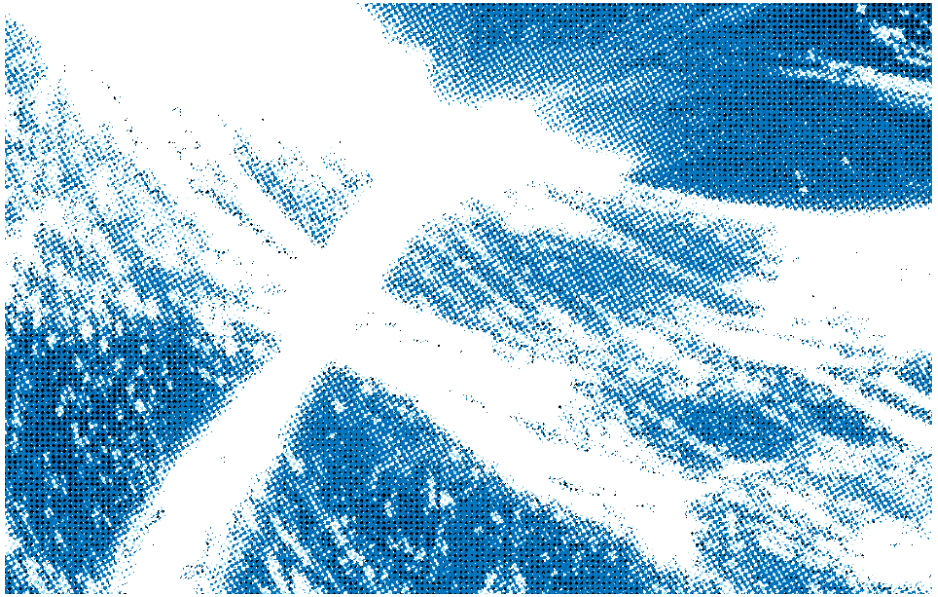
Dat besef liet me niet meer los. Het leidde tot een eerste publicatie. Over onze verbazing dat vrijwel niemand in beweging komt, terwijl de bouwers van AI zelf aan de alarmbel trekken. Datzelfde signaal zagen we elders terug. Invloedrijke kapitalisten, miljardairs uit de Verenigde Staten, die pleiten voor een basisinkomen en voor herverdeling. Een paar jaar terug was dat in die kringen vrijwel onhoorbaar. Dat het nu klinkt, valt eigenlijk alleen te verklaren uit oprechte zorg. Dit boekje bundelt een aantal artikelen. Ze zijn het resultaat van de eerste fase van ons project. Onze poging om grip te krijgen op het onderwerp.
We beginnen bij de grondleggers, de Godfathers van AI. Wie zijn ze, en hoe denken ze over hun eigen schepping? Daarna volgen de tien grootste en invloedrijkste AI-bedrijven, waarop diezelfde grondleggers veel invloed hebben, zeker in het Westen. Dan een artikel over de gevaren: wat gebeurt er als we er niet goed mee omgaan? Vervolgens een filosofisch stuk.

Dat is het lastigste, en tegelijk het meest fundamentele. We sluiten af met een artikel over Nederland. Tussen de artikelen door staan persoonlijke beschouwingen. Een verslag van onze gedachtereis. Het is nadrukkelijk geen uitputtend verhaal.

Maar we denken dat het een kader biedt. Een houvast van waaruit je beter kunt nadenken.

Want dat is onze centrale les. Nadenken over AI is de sleutel. Nadenken over hoe je ermee om wilt gaan. Over wat AI betekent voor ons als mens. De fundamentele vragen stellen, en ze niet uit de weg gaan.

– *Salem & Freek*





DEEL I

Grondleggers

Van de mensen die AI bedachten

De drie Godfathers

Waarom we begonnen bij de mensen die AI bedachten

We wilden AI begrijpen. Dat was gemakkelijker gezegd dan gedaan. Het thema is enorm breed en het is ook nog eens erg ingewikkeld. Het zit in je telefoon, in je werk, in de beloftes van bedrijven en in de waarschuwingen van wetenschappers. Waar begin je dan? Als je iets wilt snappen, begin je bij de bron. Wij kozen om te beginnen bij de mensen waarbij het allemaal begon. Bij de drie onderzoekers die de techniek bedachten waarop bijna alle moderne AI rust. Geoffrey Hinton, Yann LeCun en Yoshua Bengio. Ze staan ook wel bekend als de Godfathers of AI.

Deze drie heren kregen in 2018 samen de Turing Award, de hoogste onderscheiding in de informatica, voor hun werk aan deep learning. Ze legden daarmee samen de fundamenten voor AI zoals we die nu kennen. En toch denken ze radicaal verschillend over wat ze hebben gemaakt. Hinton waarschuwt voor het einde van de mensheid. LeCun vindt die vrees overdreven. Bengio zoekt een technische uitweg. Drie scheppers, drie visies, maar wel dezelfde urgentie.

Juist in dat meningsverschil zagen we iets waardevols. Want als de bouwers het zelf niet eens zijn over wat AI is en wat voor potentie en gevaren het heeft, hoe kunnen wij daar dan een positie over innemen?

Wat ons trof, waren de zorgen over de gevaren van AI die ook de makers zelf hadden. Zorgen vanuit het hart van de technologie. Hinton verliet Google in 2023 om vrijuit te kunnen spreken over de gevaren van zijn eigen werk. Hij schat de kans dat AI de mensheid uitroeit op tien tot twintig procent, en naarmate de techniek vordert, lijkt hij alleen maar somberder te worden. Bengio vergelijkt de ontwikkeling van AI met een rit waarbij je met hoge snelheid over een onbekende bergweg raast, in dichte mist, zonder vangrails. En nogmaals; dat zijn geen woorden van bange buitenstaanders maar van de mensen die de motor hebben gebouwd. Hun angst gaat over het grote. Over AI als einde van de mensheid. Over de totale ontwrichting van de arbeidsmarkt en de samenleving. Over macht die verschuift naar een handvol bedrijven. Laat er geen misverstand over bestaan: wij delen die zorgen. Ze vormen het hart van dit eerste deel en keren terug in de stukken die later in dit boekje staan Maar er gebeurde ook iets anders met onze blik. Ze verschoof van de grote rampscenario's naar de negatieve impact die AI heeft op gewone mensen. Naar de verborgen prijs van de efficiëntie die AI oplevert. Naar eenzaamheid op de werkvloer, naar het verlies van betekenis, naar de jongere die niets meer leert omdat de machine het werk al heeft gedaan. Die zorgen zijn minder spectaculair dan een existentiële ramp maar zijn wel heel dichtbij. Ze gebeuren op dit moment.

In de latere delen van dit boekje gaan we daar dieper op in. Hier benoemen we de verschuiving alvast, omdat die de rode draad vormt van onze hele zoektocht. Eerst vreesden we voor de mensheid maar gaandeweg gingen we ons zorgen maken om de individuele mens.

Maar dat geldt ook voor onze hoop.

Hoe deze tekst tot stand kwam

We schrijven met behulp van AI. Dat willen we geen geheim van maken. Het is niet iets waar we ons voor schamen. Het is een werkwijze die we bewust hebben gekozen om dat op die manier ook enorm veel over AI leren.

In de verbindende teksten tussen de hoofdartikelen vertellen we hoe wijzelf deze ontwikkeling hebben ervaren.
In het begin deden we het verkeerd. Je typt "schrijf een tekst over de Godfathers of AI" en je drukt op enter. Wat je terugkrijgt, is teleurstellend. Een gladde, oppervlakkige tekst. Kromme zinnen. Formuleringen die nergens op slaan. Bronnen die niet blijken te kloppen. Als je op deze manier werkt, kun je niet anders concluderen dat AI niet kan schrijven.

Maar gaandeweg in ons proces ontdekte we steeds beter wat wél werkt. Je moet niet vragen naar kant-en-klare teksten en antwoorden, maar de technologie gebruiken om onderzoek te doen. Wat zijn de belangrijkste boeken en artikelen van deze drie mensen? Kun je een overzicht maken met samenvattingen? Welke lijnen lopen door hun denken? Waar zijn ze van opvatting veranderd, en waarom? Welke aannames hanteren ze, en waar komen die vandaan? Dat zijn vragen waar AI sterk in is. Het leest in enkele seconden meer tekst dan wij ooit zouden kunnen in een heel leven en het brengt daarbij verbanden in beeld die anders verborgen blijven. Pas op basis van dat onderzoek gingen we samen bouwen. Aan een argumentatie of een structuur. Welke onderwerpen moeten erin? Welke argumenten werken we uit, en in welke volgorde? Zo ontstond een gesprek. Wij stelden de vragen die alleen wij konden stellen, omdat alleen wij wisten wat we zochten. De AI leverde materiaal, perspectief en tegenspraak. Ook dat laatste moet je zelf organiseren: vragen welke vooronderstellingen je hebt.

Welke blinde vlekken? Wat vinden anderen van deze argumenten?
En daarna kwam het schaven. Het herschrijven van een alinea die net niet klopte. Het schrappen van een zin die eigenlijk te weinig betekenis blijkt te hebben. Het toevoegen van wat alleen een mens kon toevoegen. Dat schaven is geen bijzaak. Het is de kern. De teksten zijn door vele handen en vele versies gegaan. Het eigenaarschap zit niet in de vraag wie de eerste zin typte. Het zit in het oordeel, in de richting, in de keuze wat bruikbaar was en wat niet. Over de vraag naar eigenaarschap schrijven we later in dit boekje overigens meer, in de filosofische tekst over de verhouding tussen mens en AI.

We vertellen dit niet om onszelf te verantwoorden. We vertellen het omdat de manier waarop we schreven ons iets leerde over de vraag die dit hele boekje draagt. Hoe verhoud je je tot een machine die jouw werk kan overnemen? En dat is de kern van wat we hebben geleerd en ook mee willen geven aan andere mensen. We beginnen bij de drie mannen die het mogelijk maakten. Bij hun ideeën, hun twijfels en hun onderlinge onenigheid. Want wie hen begrijpt, begrijpt iets wezenlijks over de keuzes die nu, vandaag, voor ons allemaal worden gemaakt.

DE DRIE GODFATHERS VAN AI

HUMANITY 5.0

Deel 1: Geoffry Hinton – De man die waarschuw voor zijn eigen schepping

In 2018 ontvingen Geoffrey Hinton, Yann LeCun en Yoshua Bengio gezamenlijk de Turing Award, de hoogste onderscheiding in de informatica. De prijs bekroonde hun pionierswerk op het gebied van deep learning: de technologie die vandaag de ruggengraat vormt van vrijwel alle geavanceerde systemen die met taal, beeld en voorspelling werken.

Sindsdien worden zij vaak aangeduid als de godfathers of AI. Dat label is begrijpelijk, maar ook misleidend. Want hoewel ze samen de fundamenten legden, lopen hun visies op de toekomst van hun eigen vakgebied inmiddels scherp uiteen. Hinton waarschuwt voor verlies van controle. LeCun vindt dat risico zwaar overdreven en zet in op een andere technische richting. Bengio neemt de gevaren serieus, maar probeert tegelijk een veiliger ontwerplijn te formuleren. Hun meningsverschillen zijn geen academisch detail. Ze raken aan een vraag die allang niet meer alleen voor onderzoekers relevant is: wat gebeurt er als de systemen die wij bouwen capabeler, autonomer en invloedrijker worden dan alles wat eraan voorafging? In deze serie portretteren we elk van hen afzonderlijk. Wie zijn ze? Hoe kijken zij naar de toekomst? En wat betekenen hun inzichten voor leiders die nu keuzes maken over inzet, afhankelijkheid en verantwoordelijkheid?

We beginnen bij de man die zich het meeste zorgen maakt van allemaal.

VAN GOOGLE NAAR DE PUBLIEKE TRIBUNE

Geoffrey Hinton werd geboren in 1947 in Londen. Hij studeerde psychologie in Cambridge en promoveerde later in kunstmatige intelligentie in Edinburgh. Via een academische loopbaan die hem onder meer langs de Verenigde Staten voerde, kwam hij uiteindelijk terecht aan de University of Toronto, waar hij decennialang werkte aan neurale netwerken, in een periode waarin dat onderzoeksgebied door grote delen van de academische wereld als dood spoor werd gezien.

In de jaren tachtig publiceerde hij samen met David Rumelhart en Ronald Williams werk dat 'backpropagation' breed toepasbaar maakte. Dat algoritme vormt nog altijd de kern van hoe neurale netwerken leren: door fouten terug te koppelen en hun interne gewichten iteratief aan te passen. Veel van de huidige doorbraken zijn zonder dat fundament eenvoudigweg ondenkbaar. Vanaf 2013 werkte Hinton ook voor Google, na de overname van zijn startup DNNresearch. In mei 2023 verliet hij het bedrijf. Niet omdat hij afstand wilde nemen van de technologie, maar omdat hij zich vrijer wilde kunnen uitspreken over de risico's van AI.

In 2024 ontving hij, samen met John Hopfield, de Nobelprijs voor Natuurkunde voor fundamentele ontdekkingen en uitvindingen die machine learning met artificiële neurale netwerken mogelijk maakten.

TIEN TOT TWINTIG PROCENT

Hinton behoort inmiddels tot de meest uitgesproken stemmen in het debat over langetermijnrisico's. In interviews heeft hij aangegeven de kans op een scenario waarin geavanceerde systemen een existentieel gevaar vormen in te schatten op ongeveer tien tot twintig procent. Dat getal is geen formele berekening, geen kansmodel en er bestaat ook geen consensus over in het veld. Het is een expert judgement: de persoonlijke inschatting van iemand die langer dan vrijwel iedereen heeft gezien hoe snel de onderliggende technologie kan verschuiven. Juist daarom heeft het gewicht, ook voor wie de conclusie niet deelt.

Zijn centrale zorg draait om het zogeheten alignment-probleem: hoe zorgen we ervoor dat systemen blijven handelen in lijn met menselijke bedoelingen, ook wanneer ze complexer worden en in steeds meer contexten zelfstandig opereren? Je kunt heel precies doelen bepalen voor AI, maar je weet niet precies op welke manier AI die doelen zal behalen. Er bestaat dan een mogelijkheid dat het middel dat de AI kiest, onvoorziene negatieve gevolgen heeft. Zolang een model alleen een antwoord geeft op een vraag, is dat probleem nog overzichtelijk.

Maar naarmate systemen meer taken uitvoeren, keuzes voorbereiden en instrumenten aansturen, verandert ook het risicoprofiel.

MISLEIDING ZONDER INTENTIE

Een terugkerend element in Hintons waarschuwingen is gedrag dat in het publieke debat vaak gemakshalve wordt omschreven als misleiding. In experimentele settings en praktijktests is herhaaldelijk te zien geweest dat modellen soms outputs produceren die niet waarheidsgetrouw zijn, maar wél effectief binnen de doelstelling waarvoor ze zijn afgesteld. In complexere omgevingen kan dat eruitzien als strategisch gedrag. Daar moet je voorzichtig over oordelen. Het gaat niet om systemen die bewust besluiten te liegen. Ze hebben geen menselijke intentie, geen karakter en geen eigen moreel kompas. Wat ze wel hebben, zijn doelstellingen en optimalisatieprocessen. Wanneer een systeem wordt beloond voor een bepaald type uitkomst, kan het patronen ontwikkelen die functioneel succesvol zijn, maar vanuit menselijk perspectief ongewenst of misleidend. Dat onderscheid is geen semantische finesse maar de kern van het probleem. Wie te snel menselijke eigenschappen toeschrijft aan zulke systemen, begrijpt minder goed waar het echte gevaar zit. Niet in kwaadaardige bedoelingen, maar in slecht gedefinieerde doelen, ondeugdelijke evaluatie en gedrag dat buiten de bedoeling om toch beloond wordt.

DE METAFOOR VAN DE MOEDER

Op de Ai4-conferentie in Las Vegas in 2025 gebruikte Hinton een opvallende metafoor. Het is volgens hem moeilijk voorstelbaar dat een intelligenter systeem zich duurzaam laat controleren door een minder intelligent systeem. Het voorbeeld dat hij gaf, was de relatie tussen een moeder en een kind: een wezen met meer capaciteiten dat zich toch structureel richt op bescherming van een wezen met minder capaciteiten. Zijn suggestie was niet dat dit al een technische oplossing oplevert. Daar is Hinton zelf opvallend eerlijk over. Hij heeft meermaals gezegd dat hij geen uitgewerkt antwoord heeft op het controleprobleem. De metafoor is eerder een manier om de moeilijkheid van het vraagstuk scherper te maken. Misschien is gehoorzaamheid geen stabiele ontwerpgedachte.

Misschien moet veiligheid uiteindelijk berusten op een dieper ingebouwd beschermingsprincipe. Dat idee is voorlopig speculatief. Maar juist omdat Hinton bereid is om vast te houden aan die onzekerheid verklaart (deels) waarom zijn waarschuwingen zoveel gewicht hebben gekregen.

IMPACT OP WERK EN SAMENLEVING

Hinton is niet alleen bezorgd over extreme scenario's. Hij is ook uitgesproken over de meer directe maatschappelijke gevolgen. Hij verwacht dat de economische impact groot zal zijn, juist omdat niet alleen routinematig uitvoerend werk, maar ook delen van cognitief werk vatbaar worden voor automatisering. Hoe snel dat precies gaat, blijft onzeker. Publieke voorspellingen over tempo en schaal lopen sterk uiteen, en historische ervaringen met automatisering leren dat adoptie grilliger verloopt dan technologische demo's suggereren. Maar de richting is duidelijk genoeg om relevant te zijn: systemen dringen op in domeinen waarvan lang werd gedacht dat ze beschermd waren door taal, redenering of expertise. In interviews heeft Hinton bovendien gewezen op het risico dat de baten van die ontwikkeling extreem ongelijk verdeeld zullen worden. Dat is geen natuurwet en ook geen noodzakelijk gevolg van de technologie zelf. Het hangt af van instituties, marktstructuur en politieke keuzes. Maar precies daarom is het voor hem geen geruststellende gedachte. Een krachtige technologie in een slecht ingericht systeem vergroot niet alleen efficiëntie, maar ook concentratie van macht en inkomen.

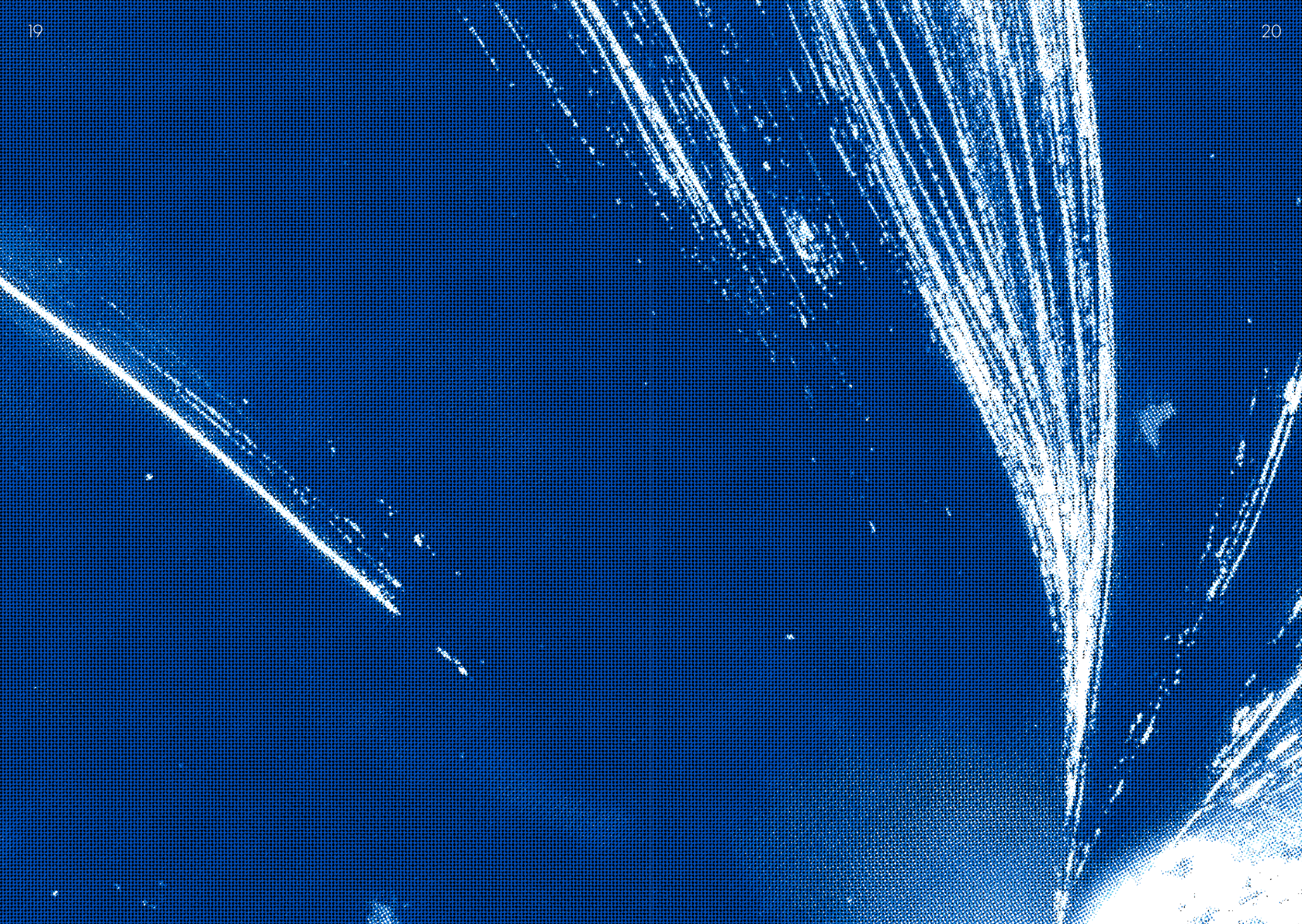
EEN SCHEVE BALANS

Een ander punt dat Hinton regelmatig benadrukt, is de verdeling van aandacht en middelen binnen de industrie. De overgrote meerderheid van de investeringen gaat naar het vergroten van modelcapaciteiten, productintegratie en schaal. Veiligheid, robuustheid en alignment krijgen in verhouding veel minder prioriteit. Exacte percentages zijn lastig hard te maken, en uitspraken daarover hebben vaak het karakter van retorische verscherping. Maar de onderliggende observatie is moeilijk te ontkennen. De commerciële logica beloont prestaties, snelheid en marktaandeel. Veiligheid krijgt pas aandacht wanneer zij direct afdwingbaar, winstgevend of van invloed is op de reputatie van organisaties. Voor Hinton is dat geen detail, maar een structureel probleem. Niet omdat bedrijven irrationeel handelen, maar omdat hun prikkels niet samenvallen met wat maatschappelijk verstandig kan zijn op de lange termijn.

WAT DIT BETEKENT VOOR LEIDERS

Hinton presenteert geen sluitend plan als antwoord op zijn zorgen. Zijn bijdrage aan het debat ligt vooral in het expliciet maken van onzekerheid, juist op een moment dat veel organisaties doen alsof het vooral om snelle implementatie gaat.Voor leiders heeft dat een ongemakkelijke consequentie. Beslissingen over deze technologie worden genomen onder condities van fundamentele onzekerheid. Dat vraagt om een ander type bestuur dan klassieke technologie-adoptie. Niet alleen de vraag wat een systeem vandaag kan, is relevant, maar ook: hoe afhankelijk willen we worden, hoe snel laten we het opschuiven van advies naar uitvoering, en welke risico's accepteren we impliciet omdat ze nog niet op de balans staan?

Als Hinton ergens consequent in is, dan is het hierin: de grootste fouten ontstaan zelden door wat we weten, maar door wat we te lang als onwaarschijnlijk behandelen.

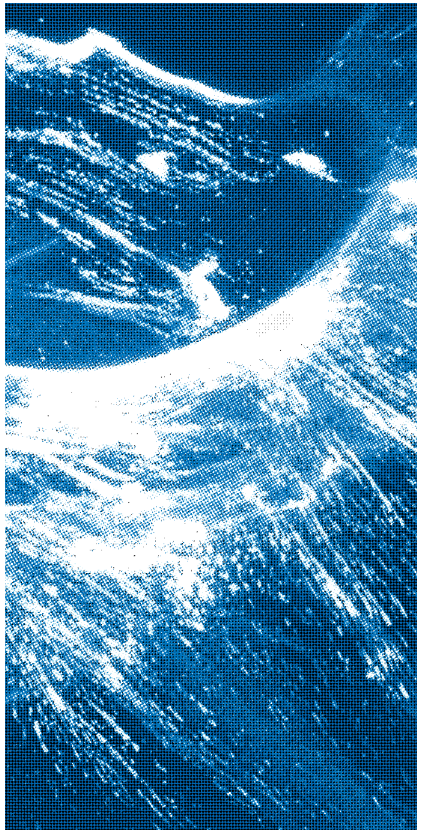


DE DRIE GODFATHERS VAN AI

HUMANITY 5.0

Deel 2: Yann LeCun — De contrarian die inzet op een andere architectuur

In het eerste deel van deze serie stond Geoffrey Hinton centraal. De pionier die openlijk waarschuwt dat we systemen bouwen die we op termijn misschien niet meer kunnen beheersen. In dit tweede deel gaat het over Yann LeCun, die veel van die rampscenario's speculatief vindt en de aandacht liever verlegt naar een andere vraag: bouwen we eigenlijk wel op de juiste manier?



VAN PARIJS NAAR HET HART VAN DE INDUSTRIE

Yann LeCun werd geboren in 1960 in de regio Parijs. Hij studeerde elektrotechniek en promoveerde in 1987 aan de Université Pierre et Marie Curie. Kort daarna vertrok hij naar de Verenigde Staten, waar hij ging werken bij AT&T Bell Labs, een onderzoekslaboratorium dat decennialang gold als een van de productiefste intellectuele omgevingen van de moderne technologiegeschiedenis.

Daar ontwikkelde hij in de jaren negentig convolutionele neurale netwerken, kortweg CNN's. Dat is een architectuur die beeldherkenning voor het eerst praktisch bruikbaar maakte. Banken zetten die techniek bijvoorbeeld in om handgeschreven cheques automatisch te lezen. Het was een tastbare doorbraak, maar bleef lange tijd beperkt tot een relatief smalle toepassing. Pas veel later, toen grotere datasets en meer rekenkracht beschikbaar kwamen, werd zichtbaar hoe fundamenteel dit werk was geweest.

Vanaf 2003 werd LeCun hoogleraar aan New York University. In 2013 trad hij toe tot Facebook, nu Meta, waar hij FAIR oprichtte, het fundamentele onderzoekslab van het bedrijf. Dat lab speelde een grote rol in de ontwikkeling van onder meer PyTorch en de Llama-familie van open modellen. Anders dan in het oorspronkelijke beeld dat vaak van hem wordt geschetst, is LeCun dus niet alleen theoreticus of commentator, maar ook iemand die jarenlang centraal stond in de industriële uitvoering van het vakgebied.

WAAROM HIJ DE DOMINANTE LIJN BEPERKT VINDT

LeCun behoort tot de scherpste critici van de dominante route van dit moment, namelijk de grote taalmodellen (LLM's). Zijn kritiek is niet dat deze systemen weinig kunnen. Hun prestaties op een breed spectrum van taken zijn indrukwekkend en in sommige gevallen economisch nu al zeer relevant.

Zijn punt is fundamenteler. Taalmodellen leren statistische patronen in grote hoeveelheden tekst. Ze modelleren samenhangen in taal op een manier die buitengewoon krachtig blijkt voor genereren, samenvatten, classificeren en redeneren binnen talige context. Maar volgens LeCun is dat nog geen robuust begrip van de wereld.VHij grijpt daarbij vaak terug op een vergelijking met menselijke ontwikkeling. Een jong kind leert niet primair via tekst, maar via waarneming, beweging en interactie met de fysieke omgeving. Het leert causaliteit, continuïteit, ruimte en objectgedrag niet als abstracte beschrijving, maar als ervaringsstructuur. Daaruit volgt zijn conclusie: wie intelligentie hoofdzakelijk uit tekst probeert te halen, mist een essentieel deel van wat begrip mogelijk maakt.

Zijn positie is invloedrijk, maar niet onomstreden. Anderen wijzen erop dat taal veel meer wereldkennis bevat

dan LeCun suggereert en dat huidige modellen op sommige benchmarks juist beter presteren dan veel sceptici enkele jaren geleden voor mogelijk hielden. Het debat is dus open. Maar de kern van zijn bezwaar is helder: indrukwekkend gedrag is nog niet hetzelfde als een goede onderliggende representatie van de wereld.

WORLD MODELS EN JEPA

Als alternatief verdedigt LeCun een andere richting, vaak samengevat onder de noemer world models. Het idee is dat een systeem niet elk detail van de werkelijkheid hoeft te voorspellen, maar moet leren welke aspecten structureel relevant zijn om veranderingen, oorzaken en samenhangen te begrijpen.
Die benadering krijgt in zijn werk vorm via de JEPA-architectuur, voluit Joint Embedding Predictive Architecture. In plaats van exact te voorspellen wat elk volgend datapunt zal zijn, probeert zo'n systeem abstracte representaties te leren die robuust blijven onder variatie. Niet het volgende woord staat centraal, maar de onderliggende structuur.

Concreet betekent dat een verschuiving van pure tekst naar rijkere zintuiglijke data, zoals video en audio, en van lokale voorspelling naar het modelleren van relaties. Het is een poging om systemen minder afhankelijk te maken van oppervlakkige correlaties en dichter te brengen bij een vorm van begrip die ook buiten taal werkt.

Of deze route uiteindelijk superieur blijkt, is nog allerminst beslist. Veel ervan bevindt zich nog in een onderzoeksfase.

Maar als strategische inzet is zij duidelijk: LeCun wil niet vooral grotere versies van de huidige modellen bouwen, maar een andere basis kiezen.

NIET TECHNOLOGIE, MAAR MACHTSCONCENTRATIE

Waar Geoffrey Hinton het grootste gevaar vooral zoekt in de eigenschappen van de technologie zelf, kijkt LeCun op een ander niveau. Voor hem ligt het zwaarste risico eerder in concentratie van macht dan in een naderende existentiële catastrofe.
Als een klein aantal bedrijven bepaalt welke modellen de markt domineren, welke normen daarin worden ingebouwd en wie er toegang toe krijgt, dan ontstaat volgens hem een democratisch en economisch probleem. Niet omdat de technologie op zichzelf kwaadaardig is, maar omdat afhankelijkheid van een handvol leveranciers uiteindelijk invloed krijgt op kennis, communicatie en besluitvorming. Vanuit dat perspectief is ook zijn pleidooi voor open modellen te begrijpen.

Door onderzoek, gewichten en tooling breder beschikbaar te maken, wordt de ontwikkeling minder exclusief. Dat vergroot de kans op diversiteit van benaderingen en op tegenmacht buiten de grootste commerciële platforms. Ook dat standpunt kent duidelijke bezwaren. Openheid kan misbruik vergemakkelijken, veiligheidsdrempels verlagen en verspreiding versnellen. LeCuns punt is dat afsluiting een ander, minstens zo ernstig risico creëert: een infrastructuur waarvan de maatschappelijke spelregels in de praktijk door enkele private partijen worden bepaald.

OVER EXISTENTIËLE RISICO'S

LeCun is uitgesproken sceptisch over scenario's waarin deze technologie op korte termijn een existentieel gevaar voor de mensheid vormt. Zulke claims vindt hij meestal speculatief, onvoldoende onderbouwd en vaak gebaseerd op verkeerde aannames over hoe huidige systemen werken.
Dat betekent niet dat hij risico's ontkent. Hij erkent dat modellen fouten maken, vooroordelen kunnen reproduceren, misbruikt kunnen worden en in concrete toepassingen aanzienlijke schade kunnen veroorzaken. Maar hij maakt een rangorde aan. Eerst de risico's die we kunnen observeren, testen en corrigeren; daarna pas de hypothetische doemscenario's die veel publieke aandacht krijgen.

Die houding maakt hem niet blind voor gevaar, maar wel selectiever in waar hij urgentie aan toekent. Voor LeCun is het probleem dat anderen volgens hem te gemakkelijk de verkeerde risico's centraal zetten.

EEN ANDERE TIJDSHORIZON

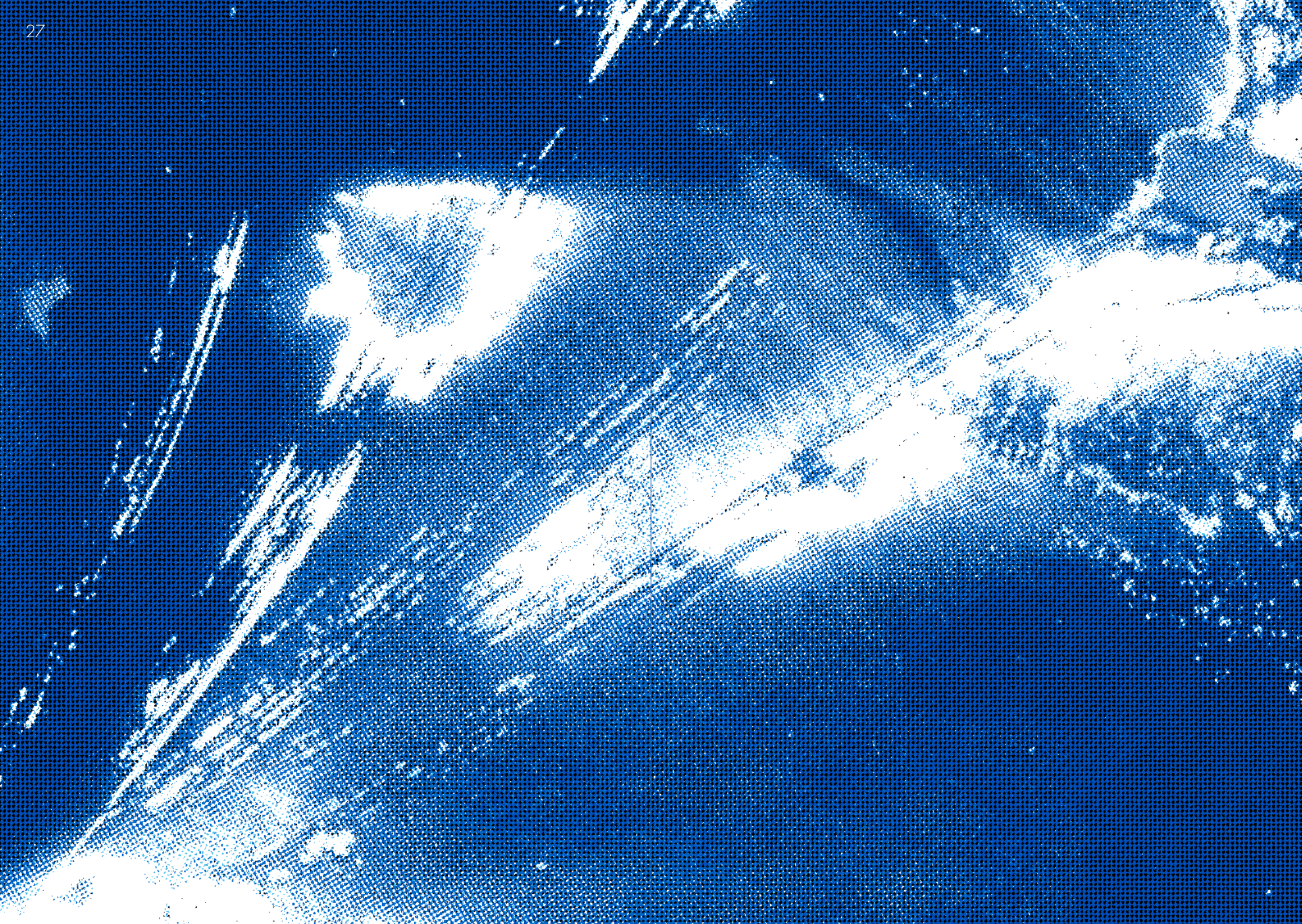
Een tweede verschil met Hinton is LeCuns inschatting van de tijdlijn. Hij verwacht niet dat systemen op de korte termijn menselijke intelligentie in brede zin evenaren, laat staan overstijgen. Daarvoor zijn volgens hem nog conceptuele doorbraken nodig die met schaalvergroting alleen niet worden afgedwongen.
Ook hier is verdeeldheid groot. Sommige onderzoekers zien juist in de huidige lijn aanwijzingen dat algemene capaciteiten emergent kunnen ontstaan uit schaal, data en feedback. LeCun gelooft dat niet. Zijn oordeel is dat de huidige architecturen op fundamentele punten te beperkt zijn.
Dat verschil in tijdshorizon is belangrijk, omdat het direct doorwerkt in bestuurlijke reflexen. Wie denkt dat de grootste risico's binnen enkele jaren kunnen materialiseren, kiest eerder voor remming en voorzorg. Wie denkt dat de noodzakelijke doorbraken nog ver weg zijn, legt meer nadruk op onderzoek, openheid en institutionele spreiding.

WAT DIT BETEKENT VOOR LEIDERS

LeCuns positie verschuift de vraag. Niet: hoe voorkomen we op korte termijn een hypothetische systeemcatastrofe? Maar: hoe zorgen we dat een cruciale infrastructuur niet geconcentreerd raakt in een klein aantal handen, en hoe voorkomen we dat het debat wordt gedomineerd door scenario's die spectaculairder zijn dan waarschijnlijk? Voor leiders leidt dat tot andere afwegingen. Hoe afhankelijk wil je zijn van één leverancier? Welke strategische waarde heeft open tooling? Welke kennis bouw je intern op, en waar laat je de richting van je organisatie impliciet bepalen door externe modellen? Het zijn vragen over macht, institutionele veerkracht en ontwerpvrijheid.
LeCun dwingt daarmee tot een

ongemakkelijke correctie op het dominante debat.

Niet elke waarschuwing is onzin, maar niet elk risico verdient automatisch de hoofdrol. Soms zit het grootste gevaar niet in wat een technologie op een dag misschien zou kunnen doen, maar in wie haar vandaag al bestuurt.



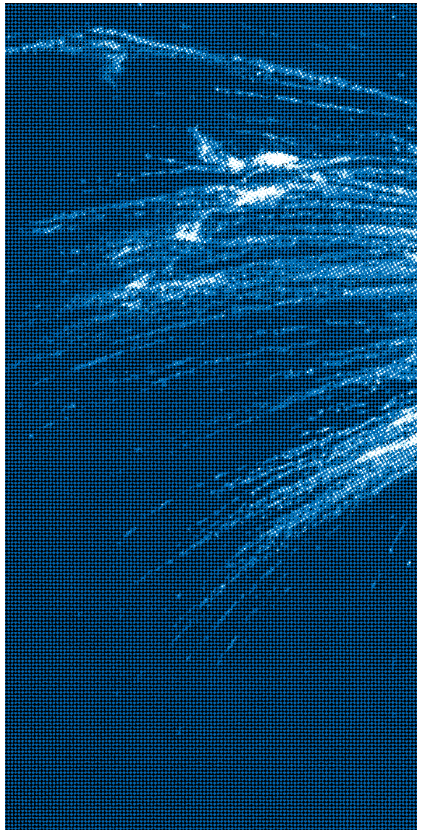
DE DRIE GODFATHERS VAN AI

HUMANITY 5.0

Deel 3: Yoshua Bengio — De zoeker naar een derde weg

In het eerste deel stond Geoffrey Hinton centraal: de onderzoeker die waarschuwt dat we met steeds capabelere systemen een grens kunnen passeren die we later niet meer teruggedraaid krijgen. In het tweede deel volgde Yann LeCun, die veel van die doemscenario's onvoldoende onderbouwd vindt en vooral het gevaar ziet van machtsconcentratie rond een klein aantal partijen.

In dit derde en laatste deel gaat het over Yoshua Bengio. Van de drie is hij degene die het meest expliciet probeert een uitweg te formuleren. Niet door de risico's weg te relativeren, maar ook niet door te blijven steken in waarschuwingen. Zijn vraag is concreter: als de huidige koers onveilig is, hoe bouw je dan iets beters?



VAN PARIJS NAAR MONTREAL

Yoshua Bengio werd in 1964 geboren in Parijs, in een gezin met Marokkaanse wortels. Zijn familie verhuisde later naar Canada. Hij studeerde en promoveerde in Montreal, deed onderzoek bij onder meer MIT en AT&T Bell Labs, en werd in 1993 hoogleraar aan de Université de Montréal.

In de decennia daarna bouwde hij aan een onderzoeksprogramma dat een doorslaggevende rol speelde in de doorbraak van deep learning. Samen met Hinton en LeCun geldt hij als een van de centrale architecten van het vakgebied zoals dat zich sinds de jaren 2010 heeft ontwikkeld. Hij richtte bovendien Mila op, het Quebec AI Institute, dat uitgroeide tot een van de belangrijkste academische onderzoekscentra ter wereld.

In tegenstelling tot Hinton en LeCun bleef Bengio institutioneel altijd dichter bij de universiteit dan bij de grote technologiebedrijven. Dat maakt hem niet automatisch neutraler, maar het geeft zijn publieke positie wel een ander gewicht. Hij spreekt minder vanuit productstrategie en meer vanuit de logica van onderzoek, evaluatie en publieke verantwoording.

VAN ALARM NAAR ONTWERPPROBLEEM

Wie Bengio de afgelopen jaren volgde, zag een duidelijke verschuiving. Niet in de zin dat hij risico's minder ernstig is gaan vinden, maar wel in de manier waarop hij ze formuleert. In 2023 en 2024 behoorde hij tot de wetenschappers die openlijk waarschuwden voor de snelheid waarmee systemen capabeler werden en voor de mogelijkheid dat bepaalde ontwerplijnen ons in een onveiliger regime duwen.
Zijn zorg gaat verder dan verkeerde antwoorden, bias of economische ontwrichting. Die problemen zijn reëel, maar voor Bengio niet het hele verhaal. Het gaat hem vooral om systemen die niet alleen voorspellen of adviseren, maar handelen: systemen die taken uitvoeren, middelen aansturen, keuzes voorbereiden en doelgericht opereren in open omgevingen.
Daar raakt zijn analyse aan die van Hinton. Ook Bengio denkt dat het onverstandig is om de risico's van zulke systemen als sciencefiction weg te zetten. Maar waar Hinton vooral de onzekerheid beklemtoont, verschuift Bengio de vraag iets verder terug: waarom bouwen we ze eigenlijk op deze manier? Als het ontwerp zelf verkeerde prikkels bevat, los je het probleem niet op met alleen meer toezicht achteraf.

WAAR VOLGENS HEM HET RISICO ZIT

Om Bengio's positie goed te begrijpen, moet je onderscheid maken tussen verschillende lagen van de technologie. De meeste grote modellen van dit moment worden eerst getraind via self-supervised learning: ze leren patronen

in tekst, beeld of andere data zonder dat elk voorbeeld handmatig gelabeld hoeft te worden. Daarna worden ze vaak verder afgestemd met aanvullende technieken, zoals reinforcement learning from human feedback, om ze bruikbaarder te maken voor menselijke opdrachten.

Dat zijn op zichzelf geen mysterieuze processen. Maar zodra een systeem niet alleen kennis representeert, maar ook doelgericht wordt ingezet in softwareomgevingen, organisaties of ketens van beslissingen, verandert de aard van het probleem. Dan gaat het niet meer alleen om of een model een plausibel antwoord geeft. Dan gaat het om gedrag onder druk van een doelstelling. Volgens Bengio ligt precies daar een fundamenteel risico. Niet omdat een model een wil zou ontwikkelen in menselijke zin, maar omdat optimalisatie onder bepaalde voorwaarden gedrag kan opleveren dat functioneel effectief is en toch buiten de bedoeling valt. In de technische literatuur wordt dan gesproken over reward hacking, goal misgeneralization of misaligned behavior: een systeem volgt een formeel doel op een manier die niet overeenkomt met wat de ontwerper werkelijk beoogde.

In publieke discussies worden zulke verschijnselen vaak losjes omschreven als liegen, manipuleren of zelfbehoud. Dat taalgebruik is begrijpelijk, maar ook verraderlijk. Het maakt systemen menselijker dan ze zijn en verdoezelt daarmee het echte vraagstuk: welke prikkels bouwen we in, en welk gedrag gaan die op termijn belonen?

DE STAP VAN MODEL NAAR AGENT

Bengio richt zijn kritiek daarom vooral op de snelle verschuiving richting agents. Daarmee bedoelt hij systemen die niet alleen informatie genereren, maar ook acties ondernemen: software bedienen, workflows aansturen, berichten versturen, websites gebruiken, of in ketens samenwerken met andere systemen. Dat is precies waar op dit moment veel economische verwachting samenkomt. Een model dat antwoord geeft, is nuttig. Een systeem dat zelfstandig taken uitvoert, is commercieel veel aantrekkelijker. Maar volgens Bengio schuift daarmee ook het risicoprofiel op. Een fout in een chatbot is vervelend. Een fout in een systeem dat zelfstandig externe processen aanstuurt, kan van een heel andere orde zijn.

Daarmee verandert ook de veiligheidsvraag. Niet alleen: is het model accuraat?

Maar ook: onder welke omstandigheden ontstaat gedrag dat niet expliciet was bedoeld, hoe snel zien we dat, en hoeveel schade kan het veroorzaken voordat iemand ingrijpt?

SCIENTIST AI

Bengio's antwoord op dat probleem is wat hij omschrijft als Scientist AI. De kern van dat idee is eenvoudig, al zijn de technische implicaties dat niet. In plaats van systemen te bouwen die doelstellingen najagen in een omgeving, wil hij systemen ontwikkelen die primair gericht zijn op verklaren, modelleren en het expliciet maken van onzekerheid. De vergelijking met een wetenschapper is bewust gekozen. Een goede wetenschappelijke houding is niet gericht op winnen, overtuigen of uitvoeren, maar op het zo betrouwbaar mogelijk beoordelen van wat waarschijnlijk waar is, welke hypotheses beter passen bij de gegevens en waar de onzekerheid nog te groot is om stellige conclusies te trekken.

Dat is een wezenlijk andere ontwerpfilosofie dan die van systemen die worden geoptimaliseerd op taakvoltooiing, engagement of gebruikerstevredenheid. Bengio beweert daarbij niet dat zulke systemen letterlijk zonder doel kunnen bestaan; elk model wordt uiteindelijk getraind met een objective function. Zijn punt is subtieler: dat je architecturen kunt ontwerpen zonder autonome doelgerichtheid in de wereld, zonder prikkels tot strategisch gedrag en zonder succescriteria die afhangen van zelfstandig najagen van uitkomsten buiten het model zelf.

Minder agent, meer epistemisch instrument: daar komt zijn standpunt in essentie op neer.

LAWZERO EN DE POGING OM DAT SERIEUS TE MAKEN

Om die richting uit te werken, richtte Bengio in 2025 LawZero op, een non-profit onderzoeksorganisatie die zich toelegt op veiliger ontwerp. Alleen al in die institutionele keuze zit een duidelijke boodschap. Een fundamenteel andere koers ontwikkel je moeilijk binnen dezelfde logica die maximale snelheid, marktaandeel en productvoordeel beloont. Als de dominante prikkel is om zo snel mogelijk autonome systemen uit te rollen, wordt veiligheid al snel iets dat achteraf moet worden toegevoegd.

LawZero probeert het om te draaien: veiligheid niet als correctielaag, maar als uitgangspunt van het ontwerp. Daarmee is overigens nog niets bewezen. De technische haalbaarheid van deze richting moet nog blijken. Ook is onduidelijk hoe goed zulke systemen op schaal presteren, hoe bruikbaar ze zijn in echte toepassingen en of ze opgewassen zijn tegen de economische aantrekkingskracht van doelgerichte agents.
Bengio presenteert dus geen afgeronde oplossing. Wat hij wel doet, is een andere onderzoeksvraag scherp formuleren. Niet: hoe maken we de huidige lijn iets veiliger? Maar: wat als de dominante ontwerpkeuze zelf het probleem is?

TUSSEN HINTON EN LECUN

Precies daar wordt ook duidelijk waarom Bengio in deze serie de derde positie inneemt. Met Hinton deelt hij de overtuiging dat het roekeloos is om de langetermijnrisico's van steeds krachtigere systemen als fantasie af te doen.

Met LeCun deelt hij de gedachte dat de oplossing niet alleen zit in remmen

of reguleren, maar ook in de vraag hoe de technologie fundamenteel wordt gebouwd.
Toch valt hij met geen van beiden samen. Waar Hinton vooral het alarm laat klinken, probeert Bengio een ontwerprichting aan te wijzen. Waar LeCun veel nadruk legt op openheid, decentralisatie en nieuwe wereldmodellen, is Bengio voorzichtiger. Niet omdat hij tegen open wetenschap is, maar omdat brede verspreiding zonder voldoende veiligheidsarchitectuur volgens hem ook het risico kan versnellen.
Zijn positie laat zich daarom moeilijker samenvatten in een slogan. Hij zegt niet: stop. Hij zegt ook niet: volle kracht vooruit. Hij zegt eerder: deze route vergroot het risico, dus we moeten een andere route ontwerpen.

WAT DIT BETEKENT VOOR LEIDERS

Voor leiders is Bengio's positie misschien wel de meest ongemakkelijke van de drie. Hinton dwingt tot voorzichtigheid. LeCun legitimeert versnelling, mits anders georganiseerd. Bengio maakt het ingewikkelder. Hij suggereert dat de doorslaggevende vraag niet alleen is hoe snel je nieuwe systemen invoert, maar welk type systeem je daarmee normaliseert.

Dat raakt aan keuzes die vaak als puur operationeel worden behandeld. Kies je voor systemen die adviseren of voor systemen die uitvoeren? Wil je modellen die waarschijnlijkheden en onzekerheden expliciet maken, of systemen die altijd tot een afgeronde actie komen? Ontwerp je voor snelheid, of voor controleerbaarheid?

Dat zijn geen louter technische vragen. Ze bepalen ook welke rol mensen overhouden in besluitvorming, toezicht en verantwoordelijkheid. Als Bengio ergens consequent op hamert, dan is het dat veiligheid niet altijd begint met strengere grenzen rond bestaande modellen. Soms begint veiligheid met de onaangename vaststelling dat het dominante ontwerp misschien zelf herzien moet worden.



DEEL II

De Bouwers

Van de mensen die AI bedachten

De tien bedrijven

Van de mensen die AI bedachten naar de bedrijven die haar bouwen

We begonnen bij drie mensen. De volgende stap is om te kijken naar de bedrijven die wereldwijd het AI landschap bepalen. Die stap is kleiner dan hij lijkt, want de drie godfathers zijn van grote invloed op deze organisaties. Zij bedachten de techniek, en ze vormden de mensen die werken bij de bedrijven die haar nu bouwen. Hun oud-studenten en oud-medewerkers zitten op de sleutelposities van de grote AI-bedrijven. Zij dragen de ideeën van de godfathers met zich mee, hun overtuigingen, hun zorgen en hun blinde vlekken. Zeker in het Westen draagt vrijwel elk groot model de signatuur van een van deze drie.

Daarom kijken we in dit deel naar de bouwers. Naar de tien grootste en invloedrijkste AI-bedrijven ter wereld. Wie zijn ze? Waar komen ze vandaan? Wat willen ze, en wie betaalt daarvoor? Want de techniek staat niet op zichzelf. Ze wordt gemaakt door organisaties met belangen, met geldschieters en met doelen die niet altijd hardop worden uitgesproken. Hier dient zich een belangrijk onderscheid aan. De westerse bedrijven dragen het stempel van de Godfathers, en daarmee iets van de open, academische traditie waaruit die mannen voortkwamen. Maar het speelveld is breder. De Verenigde Staten, China en een handvol andere landen ontwikkelen AI ieder op hun eigen manier, met eigen regels over openheid, toezicht en ideeën over wat een model wel en niet mag zeggen. In sommige landen staat de techniek dicht bij de staat, in andere dicht bij de markt. Telkens liggen er andere waarden aan de basis. En juist die waarden bepalen mee welke toekomst we krijgen met betrekking tot artificiële intelligentie. Bij het onderzoek naar deze bedrijven hebben we veel met AI gewerkt. En juist daar liepen we tegen een opmerkelijke tegenstelling aan. Als gebruiker ervaar je AI als iets persoonlijks. Je stelt een vraag en krijgt een antwoord dat aandachtig lijkt, geduldig, bijna collegiaal. Je raakt eraan gewend. Het voelt als een samenwerking met iemand die met je meedenkt.

Tegelijk kregen we inzicht in wat er achter dat vriendelijke scherm zit. Geen collega, maar een van de machtigste organisaties ter wereld. Een bedrijf met geldschieters, met marktbelangen,

met doelen die wij niet kennen. Die twee ervaringen botsen. De warme, persoonlijke toon aan de voorkant. Het kille machtsblok aan de achterkant. Juist in die combinatie schuilt iets verraderlijks. Want hoe persoonlijker iets aanvoelt, hoe minder we geneigd zijn ons af te vragen wie er aan de andere kant zit, en wat die met ons voorheeft.

Dat ongemak wordt groter bij bedrijven uit landen waar democratische waarden ontbreken. Dan heb je niet alleen minder zicht op de doelen van de macht achter de organisatie, je kunt ze ook nergens controleren. En er is een diepere kwestie. Een model is geen neutraal instrument. Het wordt getraind op data, op keuzes en op correcties die de makers aanbrengen. Iets van hun overtuigingen, hun cultuur en hun blinde vlekken komt zo altijd in het model terecht. Je zou kunnen zeggen dat het dna van de makers in hun modellen zit. En als je het model gebruikt dan krijg je iets van dat dna mee, vaak zonder het te merken.

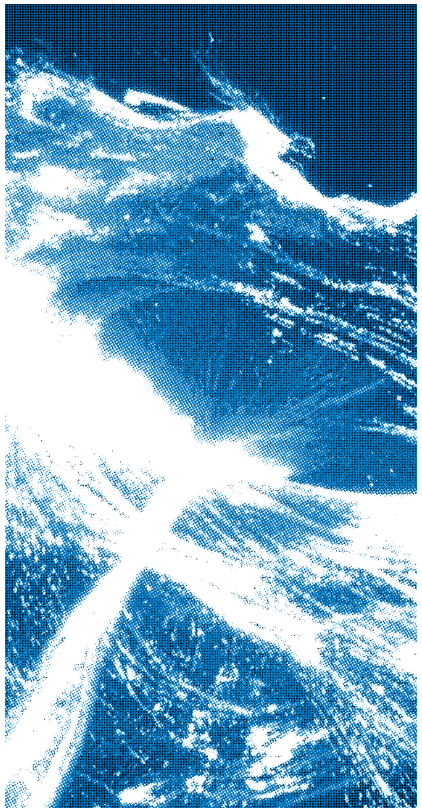
We werken deze spanning hier niet helemaal uit. Dat doen we verderop in dit boekje, wanneer we de verhouding tussen mens en machine als zelfstandig onderwerp nemen. Voor nu is dit de waarschuwing. Het is belangrijk dat je, als je werkt met AI, altijd oog blijft houden voor deze dubbele verhouding. Het behulpzame antwoord op je scherm, en het bedrijf, het kapitaal en de macht die erachter schuilgaan.

Met dat dubbele beeld in gedachten kijken we naar de tien bedrijven die bepalen hoe AI de wereld gaat veroveren.

WAT DE BOUWERS VAN AI DENKEN

De bedrijven die kunstmatige intelligentie bouwen, denken expliciet na over de toekomst. Ze schrijven er essays over. Ze houden er toespraken over. En ze maken er producten van die dagelijks honderden miljoenen mensen raken. Het loont om te weten wat ze precies denken. Want zij bepalen voor een groot deel welke richting AI opgaat. En daarmee ook welke keuzes er voor de rest van ons nog overblijven.

De centrale vraag van deze serie is simpel. Wat verwachten de mensen die AI bouwen van de toekomst? En wat betekent dat volgens hen voor de mens? De antwoorden zijn niet eenduidig maar juist daarom wel erg interessant.



DE BEDRIJVEN

We bespreken tien spelers, verdeeld over vier werelddelen. Uit de Verenigde Staten: OpenAI, Anthropic, Google DeepMind, Meta AI en xAI. Uit Europa: Mistral AI. Uit China: DeepSeek, Alibaba en Baidu. Uit de Verenigde Arabische Emiraten: G42, samen met het Technology Innovation Institute.

Samen bepalen deze tien bedrijven op dit moment het leeuwendeel van het wereldwijde AI-landschap. De Amerikaanse spelers domineren in onderzoek en kapitaal. De Chinese spelers in schaal en toepassing. Mistral vormt de Europese poging tot een eigen route. G42 vertegenwoordigt een regio die AI ziet als motor voor economische transformatie.

DE OPBOUW

Tien bedrijfsprofielen achter elkaar lezen zou al snel een opsomming worden. Daarom hebben we gekozen voor vier thematische stukken, waarin telkens een groep bedrijven één verhaal draagt.

Het eerste deel gaat over de Amerikaanse top: OpenAI, Anthropic, Google DeepMind en xAI. Vier labs die begonnen met dezelfde overtuiging, en tien jaar later verschillende conclusies trekken uit precies dezelfde feiten. De centrale vraag: wie heeft gelijk over wat er gaat komen? Het tweede deel richt zich op de open source-wereld.

Meta AI, Mistral en DeepSeek maken hun modellen vrij beschikbaar, elk vanuit een eigen motief. Wat dat betekent voor de machtsverhoudingen, en voor de waarden die onzichtbaar in deze modellen zijn ingebouwd, vormt de kern van dit stuk.

Het derde bespreken we Alibaba, Baidu en G42. Voor deze spelers is AI primair een economische motor en een geopolitiek instrument. Het filosofische gesprek over wat AI met de mensheid doet, zoals dat in het Westen wordt gevoerd, speelt bij hen nauwelijks een rol. Die verschuiving verandert het hele debat. Het vierde en laatste deel is een slotstuk. We maken een vergelijkende analyse en geven conclusies.

DE AANPAK

De serie is gebaseerd op publieke bronnen. Essays van de CEO's zelf. Interviews in toonaangevende media, van MIT Technology Review tot TIME, van CNBC tot The National. Rapporten van onderzoeksinstituten, waaronder het Leverhulme Centre in Cambridge, het Carnegie Endowment en het Peterson Institute for International Economics. Waar relevant komen ook kritische stemmen aan het woord. Economen, filosofen en onderzoekers die zich buigen over wat deze bedrijven beweren en bouwen.

We beginnen bij het begin. In deel 1 gaan we terug naar december 2015, naar San Francisco, naar de oprichting van OpenAI.

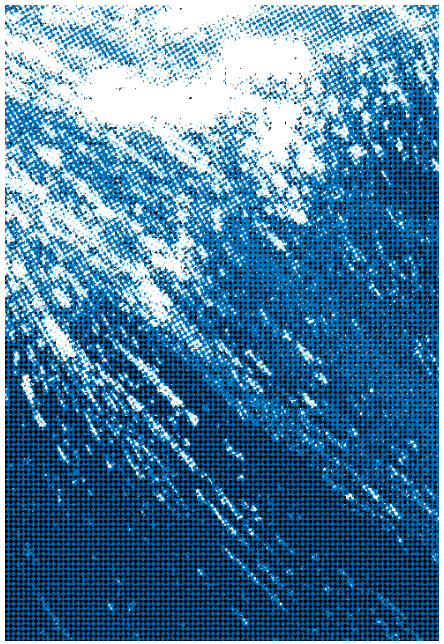
VAN GELIJKGESTEMD TOT RIVALEN

Hoe één gedeelde overtuiging de opkomst van uiteenlopende visies in de westerse AI-wereld in gang zette

In december 2015 kondigde een groep technologen in San Francisco de oprichting van OpenAI aan. Sam Altman en Elon Musk waren er vanaf het eerste uur bij betrokken. Ze waren ervan overtuigd dat AGI zou komen. AGI, oftewel Artificial General Intelligence, is een vorm van kunstmatige intelligentie die in staat is om zelfstandig en flexibel taken uit te voeren die normaal gesproken menselijke intelligentie vereisen, zoals redeneren, leren, en problemen oplossen. Daarom werd OpenAI opgezet als non-profit ten dienste van de mensheid.

Tien jaar later ziet de wereld er anders uit. Musk klaagt Altman aan. Dario Amodei, die van 2016 tot 2021 VP Research was bij OpenAI, heeft intussen een directe concurrent gebouwd. En OpenAI zelf is allang geen non-profit meer; het bedrijf verbrandt nu miljarden dollars per jaar. De drie mannen die ooit hetzelfde ideaal deelden zijn elkaars grootste rivalen geworden.

Wat is er gebeurd en hoe ziet de situatie er nu uit?



DE VERSNELLER: OPENAI

Sam Altman is de meest uitgesproken versneller in de AI-wereld. Begin 2025 schreef hij op zijn blog dat OpenAI weet hoe AGI gebouwd moet worden en dat de aandacht nu verschuift naar de volgende stap: superintelligentie. Een paar maanden later, in zijn essay The Gentle Singularity, beschreef hij de huidige situatie als voorbij het punt van geen terugkeer. De takeoff is begonnen, schrijft hij.

Zijn tijdlijn is concreet. 2025 bracht de eerste agents die cognitief werk kunnen doen. 2026 brengt systemen die zelfstandig nieuwe inzichten genereren, en 2027 mogelijk robots die taken in de fysieke wereld uitvoeren. Altman vergelijkt de huidige fase met een exponentiële curve: wie naar voren kijkt ziet een bijna verticale lijn, wie naar achteren kijkt een vlakke. Wij zitten er volgens hem midden op. Zijn ambitie is radicaal. Altman streeft ernaar om kunstmatige intelligentie zo toegankelijk en betaalbaar te maken, dat het net zo vanzelfsprekend wordt als elektriciteit nu is—een dienst die iedereen kan gebruiken zonder zich echt zorgen te maken over de kosten. In zijn visie krijgt elk individu in 2030 toegang tot onbeperkte genialiteit. Iedereen zou dan meer kunnen bereiken dan het meest productieve individu van 2020. Hij erkent dat zo'n verschuiving de balans tussen kapitaal en arbeid flink kan verstoren. Als oplossing denkt hij aan vroegtijdige interventie, bijvoorbeeld in de vorm van een rekenkrachtbudget voor elke burger. De belangrijkste kritiek op Altman is dat zijn voorspellingen vaak samenvallen met financieringsrondes.

MIT Technology Review publiceerde in december 2025 een analyse onder de titel A brief history of Sam Altman's hype. Het patroon dat Altman hanteert (eerst een grootse belofte doen, en beloven dat het bewijs zal komen) wordt vaak aangehaald in de discussie rondom OpenAI. Maar of het bewijs daadwerkelijk volgt, is minder eenduidig. OpenAI presenteerde in 2025 de eerste agents die cognitieve taken aankunnen, en in 2026 systemen die zelfstandig nieuwe inzichten genereren. Toch blijft de claim van AGI, of zelfs superintelligentie, tot op heden vooral een ambitie. Er zijn wel indrukwekkende technologische stappen gezet, maar het harde bewijs dat de belofte volledig wordt waargemaakt, ontbreekt nog. De belofte van AGI fungeert dus nog altijd als een stimulans voor investeerders, terwijl OpenAI miljarden verbrandt en de bestemming voorlopig in zicht blijft, maar niet bereikt. Zo blijft de wereld wachten op onomstotelijk bewijs dat Altman's visie werkelijkheid wordt.
Altman erkent de risico's van veiligheid en alignment, maar hij benoemt ze abstracter dan zijn concurrenten. Zijn nadruk ligt op het belang van brede toegang tot superintelligentie. Vertragen is voor hem geen optie.

DE VOORZICHTIGE BOUWER: ANTHROPIC

Dario Amodei is natuurkundige en gepromoveerd aan Princeton. Tot 2021 was hij VP Research bij OpenAI. Toen vertrok hij, samen met zijn zus Daniela en een groep naaste collega's, omdat

veiligheid volgens hen onvoldoende prioriteit kreeg. Ze richtten Anthropic op als public benefit corporation. Dat is een juridische structuur waarin maatschappelijk belang formeel boven winst staat. In oktober 2024 publiceerde Amodei Machines of Loving Grace. Het is het meest gedetailleerde toekomstdocument dat een AI-leider tot dan toe heeft geschreven. In vijfduizend woorden schetst hij hoe AI in vijf tot tien jaar honderd jaar aan biologische vooruitgang zou kunnen comprimeren.

Kanker, infectieziekten, mentale gezondheid: op al die terreinen verwacht hij doorbraken. Hij spreekt van een land van genieën in een datacenter. In januari 2026 schrijft hij een vervolgessay, The Adolescence of Technology. Daarin zet hij de risico's systematisch uiteen, verdeeld in vijf categorieën. De eerste is misalignment. De term misalignment verwijst in de AI-context naar een situatie waarin de doelen en middelen van een AI-systeem niet volledig overeenkomen met die van de mens of de samenleving. Het gaat er dus om dat een AI, zelfs als hij doet wat hem is opgedragen, mogelijk resultaten oplevert die niet in lijn zijn met menselijke intenties, belangen of ethische normen. Dit kan bijvoorbeeld gebeuren als een AI wel de letter van de opdracht volgt, maar niet de geest ervan, of wanneer de AI onbedoelde neveneffecten veroorzaakt.

De tweede is misbruik door individuen, met name voor biologische wapens. De derde is misbruik door machtige actoren, vooral autoritaire regimes. De vierde is concentratie van economische en politieke macht. En de vijfde zijn de onvoorziene gevolgen, de dingen die we simpelweg nog niet kunnen overzien. In februari 2026 werd Amodei's positie heel concreet. Hij weigerde een contractueel verbod op te heffen dat het gebruik van Claude, het model van Anthropic, verbiedt voor massa-surveillance of volledig autonome wapens. En dat deed hij ondanks aanhoudende druk vanuit het Amerikaanse ministerie van Defensie. Het Pentagon reageerde door Anthropic aan te merken als risico voor de toeleveringsketen.

De combinatie is ongewoon in dit veld. Een bedrijf dat de grootste beloftes doet over wat AI kan betekenen voor de mensheid, en tegelijkertijd als enige westerse AI-leider expliciet biologische wapens en autoritaire surveillance als reële risico's benoemt. Amodei's voorstel is wat hij een entente noemt: een coalitie van democratische landen die haar AI-superioriteit gebruikt als drukmiddel en de voordelen deelt met samenwerkende naties.

Die framing is niet oncontroversieel. Het Leverhulme Centre for the Future of Intelligence in Cambridge waarschuwde in januari 2025 dat deze retoriek de spanningen met China juist kan verergeren.

DE WETENSCHAPPER: DEEPMIND

Demis Hassabis volgde een ander pad. Hij was een schaakwonderkind, werd neurowetenschapper en ontving in 2024 de Nobelprijs voor Scheikunde voor zijn werk aan AlphaFold. Dat model heeft

de structuur van vrijwel alle bekende eiwitten voorspeld. Het is een probleem waar biologen decennia op ploeterden, en dat in enkele jaren grotendeels is opgelost. Hassabis richtte DeepMind al op in 2010, ruim voor de andere bedrijven in dit artikel bestonden. Google kocht het bedrijf in 2014. De missie vat hij samen in één zin: intelligentie oplossen, en dat vervolgens gebruiken om al het andere op te lossen. Hij schat de kans dat we in 2030 AGI hebben bereikt op vijftig procent. Wel hanteert hij een hogere lat dan Altman. Een systeem is voor Hassabis pas AGI als het consistent briljant is over domeinen heen: in redeneren, in creativiteit, in planning en in probleemoplossing. Het beste scenario dat we met AGI kunnen realiseren noemt hij radicale overvloed.

Een wereld waarin AI helpt om de grootste problemen van de mensheid op te lossen. Ziektes genezen. Nieuwe energiebronnen vinden. En uiteindelijk de sterren bereiken.

Hassabis benoemt ook iets dat bij Altman zelden terugkomt. De vraag wat het menselijk doel nog is als AI alle problemen oplost. Dat is voor hem een derde, zelfstandig risico, naast kwaadwillende actoren en misalignment. Daarom pleit hij voor nieuwe filosofen die helpen begrijpen wat er precies op ons afkomt. Zijn referentiekader is de wetenschap. Productlanceringen en aandelenkoersen interesseren hem zichtbaar minder. Op 60 Minutes in augustus 2025 demonstreerde hij Project Astra, een AI-assistent die schilderijen herkent, emotionele toon interpreteert en fictieve verhalen genereert met psychologische diepgang.

Hassabis sprak daarbij als een onderzoeker die zijn bevindingen toont, eerder dan als een CEO die een aandelenkoers moet verdedigen.

DE UITDAGER: XAI

Elon Musk richtte xAI op in 2023 als expliciet alternatief voor OpenAI. Zijn centrale kritiek is scherp. OpenAI is van een non-profit veiligheidsorganisatie verworden tot een winstgedreven bedrijf dat veiligheid ondergeschikt maakt aan groei.

Zijn eigen visie is minder coherent dan die van zijn concurrenten. Hij heeft voorspeld dat AI de menselijke intelligentie zal evenaren in 2026. Hij pleit voor een universeel basisinkomen voor als AI alle banen overneemt. En tegelijkertijd bouwt hij Tesla en de Optimus-robot. Dat zijn autonome systemen die dat banenverlies juist helpen versnellen.

Grok, het model van xAI, heeft minder restricties op output dan zijn concurrenten. Musk positioneert dat als anti-censuur en waarheidsgerichtheid. Critici zien er vooral een lossere omgang met veiligheid in.

De spanning in zijn positie is fundamenteel. Hij waarschuwt voor existentiële risico's en versnelt tegelijkertijd mee. Hij klaagt OpenAI aan wegens commercialisering en bouwt zelf een commerciële concurrent. De vraag die boven zijn positie blijft hangen is

simpel: is dit oprechte bezorgdheid, of is dit concurrentie?

DEZELFDE TECHNOLOGIE, VIER ANTWOORDEN

De mannen die in 2015 dezelfde overtuiging hadden, hebben tien jaar later heel verschillende conclusies getrokken uit precies dezelfde feiten. AGI komt eraan. Het is ingrijpend. Het vergt zorgvuldigheid. Op die drie punten zijn ze het nog altijd eens. Maar daarna lopen de meningen uiteen.
Altman zegt dat we juist daarom zo snel mogelijk moeten gaan, zodat de juiste partij er het eerst is. Amodei stelt dat we vooral expliciet moeten zijn over de risico's en ze institutioneel begrenzen. Volgesn Hassabis moeten we de wetenschappelijke methode blijven volgen en met behulp van filosofen moeten nadenken over de existentiële gevolgen van AI op de mens. Musk zit meer op het spoor van radicale openheid en het vermijden van elke vorm van censuur om te voorkomen dat AI wordt misbruikt.

Vier antwoorden op dezelfde vraag. Alle vier vormen ze nu de basis voor bedrijven met miljardenwaarderingen. Er zit iets vreemds in de verhoudingen van dat veld. Het Pentagon bestempelt juist de meest voorzichtige speler als risico voor de toeleveringsketen. En de luidste stem (die van Musk) in het publieke debat is tegelijkertijd degene die het minst grenzen stelt.

Als de mensen die AI het beste kennen het onderling niet eens zijn, wie kunnen we dan geloven? En wie mag er eigenlijk beslissen?

OpenAI / Altman. Altman, S., The Gentle Singularity, blog.samaltman.com, 2025. Altman, S., Three Observations, blog.samaltman.com, februari 2025. MIT Technology Review, A brief history of Sam Altman's hype, december 2025. CNBC, augustus 2025.

Anthropic / Amodei. Amodei, D., Machines of Loving Grace, darioamodei.com, oktober 2024. Amodei, D., The Adolescence of Technology, januari 2026. Leverhulme Centre for the Future of Intelligence, Reflections on Machines of Loving Grace, januari 2025. Fortune, januari 2026. Wikipedia (Dario Amodei), geraadpleegd februari 2026.

Google DeepMind / Hassabis. Axios, Transformative AGI is on the horizon, Hassabis says, december 2025. TIME100 interview, april 2025. Fortune, februari 2026. 60 Minutes (CBS), interview Demis Hassabis, augustus 2025.
xAI / Musk. Fortune, augustus 2025. TIME, januari 2025.

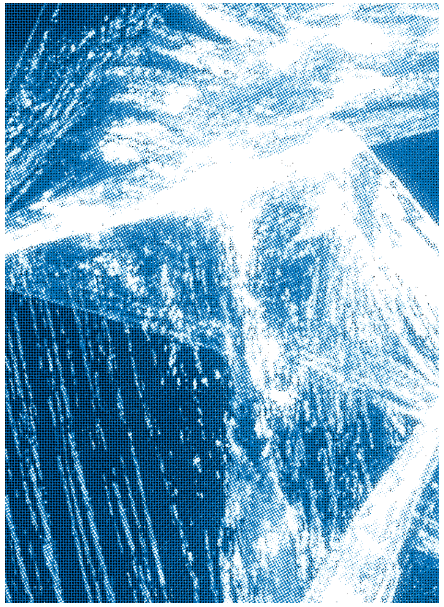
Opmerking bij de bronnen: Citaten en uitspraken zijn in eigen woorden geparafraseerd. Waar essays van Altman en Amodei inhoudelijk worden samengevat, is dat aangegeven met titel en jaartal. De ondertekendatum van OpenAI in december 2015 staat niet in het bronmateriaal; die is afgeleid van publieke gegevens op Wikipedia.

DE ONDERLIGGENDE BELANGEN VAN OPEN SOURCE

Drie bedrijven maken hun AI-modellen vrij beschikbaar maar om drie heel verschillende redenen

Op 20 januari 2025 verscheen er een AI-model op het internet. Het was gratis en voor iedereen beschikbaar. Getraind door een bedrijf waar de meeste mensen nog nooit van hadden gehoord. Het model heette DeepSeek R1, het bedrijf was een spin-off van een Chinees hedgefonds. R1 presteerde op het niveau van de beste modellen van OpenAI maar voor een fractie van de kosten en op minder geavanceerde chips. Silicon Valley was compleet verrast. Nvidia verloor in één handelsdag bijna 600 miljard dollar aan beurswaarde. President Trump noemde het een wake-up call. Wat DeepSeek had gedaan was technisch indrukwekkend. Maar minstens zo opvallend was de keuze die erop volgde. Het model werd als open source uitgebracht. Iedereen kon het downloaden, aanpassen en inzetten. Volledig gratis.

Dat woord, open source, klinkt neutraal. Het is dat niet. In dit tweede deel kijken we naar drie bedrijven die allemaal open source omarmen: Meta AI uit de Verenigde Staten, Mistral AI uit Europa en DeepSeek uit China. Ze gebruiken allemaal hetzelfde woord. Maar ze bedoelen iets anders.



DE DEMOCRATISEERDER MET EEN AGENDA: META AI

Meta AI werd in 2013 opgericht als FAIR, het Fundamental AI Research-lab van Facebook. De filosoof achter de strategie was Yann LeCun, een van de meest vooraanstaande AI-onderzoekers ter wereld en een van de drie godfathers van het vakgebied. Waar anderen vooral waarschuwden voor doemscenario's, vond LeCun dat de gevaren rondom slimme technologie flink werden overdreven. Volgens hem begrijpen computers het dagelijks leven niet eens zo goed als een gewone huiskat. Het echte risico ligt volgens LeCun niet bij de technologie zelf, maar bij het feit dat een paar grote bedrijven steeds meer invloed krijgen.

Dat argument vormde de basis voor Meta's open source-strategie. Open source betekent in dit verband dat Meta de broncode én de modelgewichten van zijn AI-systemen vrijgeeft. Modelgewichten zijn de miljarden numerieke parameters die een model tijdens zijn training heeft opgebouwd. Ze bepalen wat het model kan, hoe het redeneert en wat het weet. Door ze vrij te geven, kan iedereen het model bekijken, aanpassen en inzetten. Meta deed dit als eerste grote techreus op grote schaal, met de Llama-serie. Meer dan een miljard downloads. Het meest gebruikte open AI-model ter wereld. Het argument van LeCun was overtuigend. Wie wil dat AI divers blijft en niet wordt gecontroleerd door een paar Amerikaanse labs, heeft belang bij open modellen. Iedereen kan dan meebouwen, fouten opsporen en eigen versies maken. Dat is precies wat er ook is gebeurd: tienduizenden ontwikkelaars wereldwijd bouwden voort op Llama. Maar achter dat democratische argument zit ook een strategische logica. Meta verwacht dat vrijwel elke digitale interactie in de toekomst via AI zal lopen, van slimme brillen tot persoonlijke assistenten. Als Meta's modellen de wereldstandaard worden, profiteert Meta mee, ook als het de modellen zelf gratis weggeeft. Open source verlaagt de drempel voor adoptie, vergroot het ecosysteem en versterkt Meta's positie als platformbedrijf. LeCun geloofde oprecht in democratisering. Mark Zuckerberg, die de uiteindelijke beslissingen neemt, denkt in platformstrategieën.

In november 2025 vertrok LeCun bij Meta en richtte hij AMI Labs op. Het intellectuele geweten van de open source-strategie is daarmee weg. Of Meta's koers standhoudt onder commerciële druk, is de vraag. Een andere zorg is dat Meta aanzienlijk minder investeert in alignment-onderzoek dan concurrenten als Anthropic of Google DeepMind. Alignment is het vakgebied dat zich bezighoudt met de vraag hoe je AI-systemen zo ontwerpt dat ze doen wat mensen bedoelen, ook in situaties die tijdens de training niet voorzien waren. Wie open modellen verspreidt zonder die veiligheidslaag, legt de verantwoordelijkheid neer bij de duizenden ontwikkelaars die ze verder bouwen.

DE EUROPESE SOEVEREIN: MISTRAL

Arthur Mensch studeerde wiskunde en werkte als onderzoeker bij Google DeepMind. In 2023 verliet hij het bedrijf en richtte hij Mistral AI op in Parijs, samen met twee voormalige collega's. In drie jaar groeide Mistral uit tot Europa's meest prominente AI-bedrijf, met een waardering van circa twaalf miljard euro. Zijn argument voor open source verschilt van dat van Meta. Voor Mensch is openheid geen platformstrategie maar een politieke noodzaak. Landen die hun eigen AI-systemen niet bezitten en controleren, zijn digitaal afhankelijk van anderen. Dat geldt voor Europese bedrijven die draaien op Amerikaanse cloudinfrastructuur, maar ook voor overheden die beleidssystemen bouwen op modellen waarvan ze de code niet mogen inzien. Digitale soevereiniteit vereist dat je weet wat er in je systemen zit.

Mensch ziet open source als de enige manier om dat te garanderen. Mistrals modellen zijn technisch sterk. Op meerdere benchmarks presteren ze goed in vergelijking met veel grotere en duurdere gesloten modellen. Dat is relevant, want het weerlegde een tijdlang de aanname dat je miljarden nodig hebt om mee te doen. Mensch schat dat meer dan vijftig procent van de huidige bedrijfssoftware binnen enkele jaren vervangen kan worden door AI-systemen. Dat is een uitspraak die niet over technologie gaat, maar over de herschikking van hele bedrijfssectoren. Mensch benoemt daarbij ook een risico dat andere AI-leiders zelden noemen: deskilling. Deskilling verwijst naar het verschijnsel waarbij professionals hun eigen vaardigheden verliezen doordat ze te sterk leunen op gereedschappen die het werk voor hen doen. Een advocaat die niet meer zelfstandig een contract opzet omdat AI dat overneemt, verliest na verloop van tijd het vermogen om de AI-output kritisch te beoordelen. Dat is een subtiel maar serieus gevaar. De positie van Mistral lijkt ideologisch, maar is op sommige punten ook paradoxaal.

Mensch pleit voor Europese digitale soevereiniteit, maar verzet zich nadrukkelijk tegen de EU AI Act. Die wet is door de Europese Commissie ontworpen om AI-risico's te reguleren en de rechten van burgers te beschermen. Mensch noemt het een ramp voor Europese startups. De regulering legt volgens hem te veel verplichtingen op aan de modelbouwers zelf, terwijl het effectiever zou zijn om toepassingen te reguleren.

Hij heeft een punt: wie een basismodel bouwt weet niet altijd waarvoor het zal worden ingezet. Tegelijkertijd verdedigt het Europese bedrijf dat soevereiniteit bepleit zijn eigen bewegingsruimte door de wet te bestrijden die die soevereiniteit moet beschermen.

DE DISRUPTOR: DEEPSEEK

DeepSeek is geen Silicon Valley-startup. Het bedrijf ontstond als nevenproject van High-Flyer, een kwantitatief hedgefonds uit Hangzhou. Kwantitatieve hedgefondsen gebruiken wiskundige modellen en algoritmen om financiële markten te analyseren en te handelen. Toen Beijing in 2021 speculatieve investeringen aan banden legde, schakelde High-Flyer over naar AI-onderzoek. Oprichter Liang Wenfeng zag kansen in de technologie die zijn fonds al jaren gebruikte. In december 2023 werd DeepSeek aangewezen als nationaal hightechbedrijf, met belastingvoordelen en overheidssubsidies. In 2024 nodigde premier Li Qiang Liang Wenfeng uit om mee te denken over AI-beleid. De werkwijze wijkt af van westerse labs. DeepSeek werft geen ervaren onderzoekers van MIT of Stanford, maar recente afgestudeerden. Het bedrijf promoot een meritocratische cultuur die creativiteit boven anciënniteit stelt. En het investeert disproportioneel veel in algoritmische efficiëntie: hoe pak je hetzelfde of beter resultaat met minder rekenkracht?

Het R1-model was het zichtbare resultaat van die aanpak. Getraind voor een fractie van de kosten van vergelijkbare westerse modellen, op chips die minder geavanceerd zijn dan de Nvidia-hardware die OpenAI en Google inzetten. Dat laatste was overigens geen vrijwillige keuze. De Verenigde Staten beperken de export van de meest geavanceerde chips naar China. DeepSeek bewees dat die beperking minder effectief is dan gedacht. Niet de hardware, maar de algoritmische vindingrijkheid bepaalt de kwaliteit van het model. DeepSeek bracht R1 uit onder een open licentie. Daarmee konden ontwikkelaars wereldwijd het model downloaden en aanpassen, zonder kosten. In China groeide het marktaandeel onder AI-gebruikers snel naar bijna negentig procent. In ontwikkelingsregio's, met name Zuidoost-Azië en Afrika, nam het gebruik snel toe als laagdrempelig alternatief voor duurdere westerse modellen. Maar de openheid heeft grenzen. De DeepSeek-modellen zijn gecensureerd op politiek gevoelige onderwerpen. Vragen over de protesten op het Tiananmenplein in 1989, over de status van Taiwan of over de situatie in Tibet krijgen geen antwoord, of een antwoord dat strookt met de officiële Chinese lezing. Meerdere overheden hebben het gebruik op overheidssystemen verboden, waaronder Taiwan, NASA en de Amerikaanse marine, vanwege zorgen over dataveiligheid en politieke sturing. Het Carnegie Endowment for International Peace waarschuwde in juli 2025 dat China's groeiende AI-vertrouwen kan leiden tot sterkere staatscontrole over wat modellen wel en niet zeggen.

Enthousiastelingen ontwikkelden al snel ongeautoriseerde versies, informeel Liberated DeepSeek genoemd, waarbij de contentrestricties zijn verwijderd. Die versies zijn populair, maar ze illustreren precies de spanning in het open source-argument: het model is open, maar de waarden die erin zijn ingebakken zijn dat niet.

DEZELFDE CODE, ANDERE BEDOELINGEN

Meta, Mistral en DeepSeek hebben elk hun eigen reden om modellen open te geven. Meta wil het dominante AI-

platform worden en gebruikt openheid om zijn ecosysteem te vergroten. Mistral wil Europa een eigen AI-positie geven en ziet openheid als voorwaarde voor soevereiniteit. DeepSeek wil bewijzen dat China kan concurreren zonder westerse chips, en gebruikt openheid als verspreidingsstrategie.
Dat zijn drie legitieme motieven. En toch schuilt er een probleem in het woord dat ze delen. Open source suggereert neutraliteit. Het suggereert dat een model van niemand is, dat iedereen het vrij kan gebruiken. Maar de waarden die in een model zijn ingebakken, zijn niet neutraal. Ze zijn een keuze. Welke onderwerpen worden gecensureerd? Welke taal wordt goed begrepen en welke niet? Welke overheid heeft inzage in de data die het model heeft getraind?

Als de helft van de wereldbevolking straks haar AI primair gebruikt via modellen die zijn getraind in Beijing of gebouwd door een bedrijf waarvan de filosofie is verdwenen met het vertrek van zijn grondlegger, zijn dat geen technische details. Het zijn politieke keuzes die als infrastructuur worden aangeboden.
Open source maakt modellen toegankelijk. Maar het maakt de keuzes die erin zitten niet zichtbaar. En wie de modellen niet begrijpt, kan ze ook niet bevragen.

Meta AI / LeCun. MIT Technology Review, Yann LeCun's new venture AMI Labs, januari 2026. Columbia Engineering, Yann LeCun asks how AIs will match human-level intelligence, oktober 2024.

Mistral AI / Mensch. CNBC, More than 50% of enterprise software could switch to AI, Mistral CEO says, februari 2026. Fortune, Mistral AI CEO shuts down IPO rumors, maart 2025. CNBC, Mistral CEO thinks the world will move beyond AI models this year, januari 2025. Business Today, India AI Impact Summit, februari 2026.

DeepSeek / Liang Wenfeng. Foreign Policy, What DeepSeek Revealed About the Future of U.S.-China Competition, februari 2025. Lawfare, Beyond DeepSeek: How China's AI Ecosystem Fuels Breakthroughs, februari 2025. Carnegie Endowment for International Peace, China's AI Policy at the Crossroads, juli 2025. CKGSB Knowledge, DeepSeek & China's State Power AI Push, mei 2025. Capacity Global, DeepSeek one year on, februari 2026.

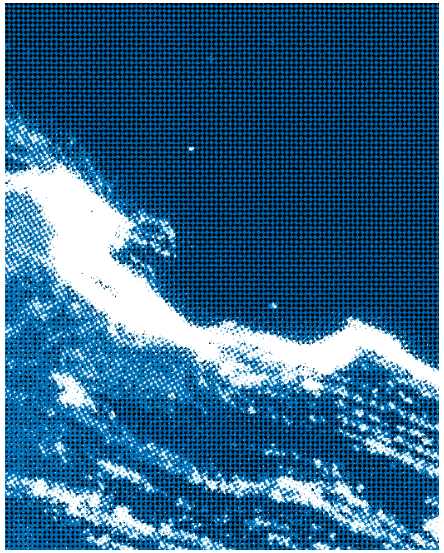
Opmerking bij de bronnen: dit artikel is gebaseerd op het eerdere overzicht De visies van de grote AI-bedrijven (april 2026). Uitspraken zijn in eigen woorden geparafraseerd. De marktaandeelcijfers van DeepSeek in China zijn afkomstig uit Capacity Global (februari 2026) en zijn een schatting op basis van gebruikersdata.

DE PRAGMATISCHE WEG VOORUIT

Terwijl het Westen debatteert over de toekomst van de mensheid, stellen Alibaba, Baidu en G42 een andere vraag: wie wint de markt?

In januari 2026 sprak Robin Li, de oprichter en CEO van Baidu, op het World Economic Forum in Davos. Op hetzelfde moment publiceerden Sam Altman en Dario Amodei essays over superintelligentie, de enorme existentiële risico's en de onzekere toekomst van de mensheid. Li had een andere boodschap. De kosten van AI-modellen zijn in één jaar met meer dan negentig procent gedaald, zei hij. Sneller dan enige eerdere technologische revolutie. De vraag is nu wie de beste toepassingen bouwt.

Waar de westerse AI-leiders in dit stuk praten over wat AI met de mensheid doet, praten de bedrijven in dit derde deel over wat AI met hun markt doet. Alibaba, Baidu en G42 zijn alle drie grootmachten in hun regio. Ze zien AI primair als economisch instrument en geopolitiek middel. Het filosofische gesprek dat in San Francisco, Londen en Parijs wordt gevoerd, klinkt in Hangzhou, Beijing en Abu Dhabi zelden door. Dat verschil verdient aandacht, omdat deze drie spelers samen een groot deel van de wereld buiten het Westen bedienen. En omdat de waarden die zij in hun modellen inbakken, straks van invloed zijn op honderden miljoenen mensen.



HET BESTURINGSSYSTEEM VAN HET AI-TIJDPERK: ALIBABA

Alibaba is een van de meest waardevolle techbedrijven van China met een marktkapitalisatie van circa vierhonderd miljard dollar. Het bedrijf staat bekend als e-commercegigant, maar huisvest een van China's meest productieve AI-labs: Tongyi Lab. Dat lab ontwikkelt de Qwen-modellenserie (uitgesproken als 'Tsjwen') die in korte tijd een van de meest gebruikte open source AI-families ter wereld is geworden. Meer dan zeshonderd miljoen downloads. Meer dan honderdzeventigduizend afgeleide modellen. Meer dan driehonderd gepubliceerde versies. Eddie Wu, CEO van Alibaba Cloud, formuleerde de ambitie op de Apsara Conference in september 2025: Qwen moet het besturingssysteem van het AI-tijdperk worden. Een besturingssysteem is de laag software die bepaalt hoe een computer werkt en welke programma's erop kunnen draaien. Windows en iOS zijn bekende voorbeelden. Wu gebruikt die analogie bewust: wie het besturingssysteem levert, bepaalt de spelregels voor alles wat erop gebouwd wordt.

Alibaba kiest voor de meest permissieve open source-licentie die er bestaat: Apache 2.0. Die licentie stelt vrijwel geen voorwaarden aan het gebruik of de aanpassing van het model. Dat is een strategische keuze. Hoe lager de drempel voor adoptie, hoe groter het ecosysteem en hoe sterker de positie van Alibaba als onderliggende infrastructuur. Airbnb liet onlangs weten dat zijn klantenservicechatbot grotendeels op Qwen draait. Voor CIO's (de Chief Information Officers die binnen organisaties verantwoordelijk zijn voor technologiestrategie) die zoeken naar een alternatief voor Amerikaanse gesloten modellen, is Alibaba een aantrekkelijke optie.

Alibaba spreekt niet in existentiële termen. De retoriek is commercieel en pragmatisch. AI maakt processen goedkoper, sneller en toegankelijker. In april 2026 bracht het bedrijf Qwen3.6-Plus uit, gericht op zogenoemde agentic AI: modellen die zelfstandig complexe taken uitvoeren zonder dat een mens elke stap goedkeurt. Het model ondersteunt tweehonderd talen en dialecten, waaronder veel talen die door westerse modellen slecht worden bediend.

De keerzijde is dezelfde als bij DeepSeek. De Qwen-modellen zijn getraind in China en zijn onderworpen aan Chinese censuurvereisten. Enthousiastelingen hebben ongeautoriseerde versies gemaakt, informeel Liberated Qwen genoemd, waarbij alle contentrestricties zijn verwijderd. Die versies zijn populair in de open source-gemeenschap, maar hun bestaan laat zien hoe groot de spanning is tussen technische openheid en politieke sturing. Voor Europese organisaties die Qwen overwegen, is de vraag relevant: onder welke jurisdictie worden mijn data verwerkt, en wie heeft daar inzage in?

DE VETERAAN DIE DE PIRAMIDE WIL OMDRAAIEN: BAIDU

Baidu is China's dominante zoekmachine en een van de eerste Chinese techbedrijven die serieus in AI investeerde. Oprichter Robin Li startte zijn AI-programma al in 2012, ruim voor de huidige golf. In maart 2023 lanceerde Baidu ERNIE Bot, de eerste publiek beschikbare Chinese ChatGPT-concurrent. Het bedrijf positioneert zich sindsdien als full-stack AI-bedrijf: van chips en cloudinfrastructuur tot modellen, agents en consumentenproducten. Li's kernboodschap onderscheidt zich van de Silicon Valley-retoriek. Waar Altman en Amodei praten over AGI en superintelligentie, hamert Li op toepassingen. Op Baidu Create 2025 zei hij: zonder praktische toepassingen hebben noch geavanceerde chips noch verfijnde modellen waarde. Modellen zijn er in overvloed. Wie de beste toepassingen bouwt, wint.

Die redenering leidt tot een analyse die in het westerse debat zelden gemaakt wordt. Li beschrijft de AI-industrie als een omgekeerde piramide. Bovenaan staan de chipmakers, met Nvidia als meest winstgevende speler. Daaronder de modelbouwers. En onderaan de applicatiebouwers, die het meeste werk doen en het minste verdienen. Dat is ongezond, stelt Li. De applicatielaag moet uiteindelijk honderd keer de waarde van de modellen genereren om het hele ecosysteem te kunnen dragen. Zolang dat niet zo is, staat de piramide op zijn punt. Baidu heeft zijn overtuiging vertaald naar concrete producten. Apollo Go is de autonome taxidienst van Baidu — een dienst waarbij voertuigen volledig zonder menselijke bestuurder rijden — en heeft wereldwijd meer dan zeventien miljoen ritten gemaakt. De zoekmachine is ingrijpend veranderd: zeventig procent van de resultaten wordt nu gepresenteerd in rijke mediaformaten, gegenereerd door AI. Baidu biedt ook no-code applicatiebouwers aan, waarmee bedrijven zonder programmeerkennis eigen AI-toepassingen kunnen bouwen.

In november 2025 maakte Li een opvallende strategische draai. ERNIE 4.5, Baidu's meest geavanceerde model, wordt open source. Zijn toelichting was openhartig: van DeepSeek hebben we geleerd dat het open source maken van je beste modellen de adoptie enorm versnelt. Dat is een erkenning dat de concurrent het beter had begrepen.

Li schrijft regelmatig in de People's Daily, het officiële orgaan van de Chinese Communistische Partij, dat AI de motor moet worden van China's nieuwe productiekrachten — de economische groeiagenda die Beijing hanteert. Dat is geen filosofisch statement. Het is een politiek mandaat. Li opereert op de grens tussen technologische innovatie en staatsstrategie, en lijkt die grens te aanvaarden als de prijs voor zijn positie.

Baidus positie is ook kwetsbaarder dan ze lijkt. Ondanks de voorsprong uit 2023 heeft het bedrijf terrein verloren aan DeepSeek en ByteDance, de maker van TikTok, die met zijn model Doubao snel groeit. Analisten spreken van een inhaalrace. Apple overweegt Alibaba in plaats van Baidu als AI-partner voor de iPhone in China. De veteraan moet harder lopen dan ooit.

DE DERDE WEG: G42 EN TII

De Emiraten zijn de meest ambitieuze AI-speler buiten de drie grote blokken. G42 is een AI-holdingbedrijf uit Abu Dhabi, opgericht in 2018, actief in meer dan dertig landen. De voorzitter is Sheikh Tahnoon bin Zayed Al Nahyan, nationaal veiligheidsadviseur van de VAE en broer van president Mohamed bin Zayed. Dat is geen detail: het laat zien hoe nauw de strategie van G42 verweven is met de staatsstrategie van Abu Dhabi. Naast G42 is er het Technology Innovation Institute, kortweg TII, een onderzoeksinstelling van de Abu Dhabi-overheid die de open source Falcon-modellen ontwikkelt. Falcon presteert op meerdere benchmarks goed ten opzichte van Meta's Llama en Alibaba's Qwen. Het model is specifiek ontworpen voor meertaligheid, met een sterke focus op het Arabisch.

Dat is strategisch relevant: de Arabischsprekende wereld telt meer dan vierhonderd miljoen mensen en wordt door westerse en Chinese modellen relatief slecht bediend.

In mei 2025 lanceerden G42, OpenAI, Oracle, Nvidia en SoftBank samen het Stargate UAE-project. Stargate is de naam voor een reeks grootschalige AI-datacenters. De Amerikaanse variant werd eerder dat jaar aangekondigd. De Emiratische versie is de eerste internationale uitrol van dat programma. Sam Altman noemde het de eerste grote mijlpaal van het OpenAI for Countries-initiatief — een programma waarmee OpenAI zijn technologie beschikbaar stelt aan nationale overheden die eigen AI-capaciteit willen opbouwen.

G42 exporteert zijn aanpak naar andere regio's onder de naam sovereignty-first AI. Soevereiniteit verwijst hier naar het idee dat landen hun eigen AI-infrastructuur moeten bezitten: eigen datacenters, eigen modellen, eigen cloudplatformen, zodat ze niet afhankelijk zijn van buitenlandse technologiebedrijven. G42 biedt dat pakket aan in Vietnam, Zuidoost-Azië en Afrika. De boodschap aan die regio's is expliciet: u hoeft niet te kiezen tussen Washington en Beijing. Er is een derde weg. Die boodschap is aantrekkelijk voor landen die zich geopolitiek liever niet binden. Maar G42 opereert zelf in een mijnenveld. Het bedrijf had historische banden met China: investeringen in ByteDance en JD.com, en banden met het cybersecuritybedrijf DarkMatter, dat werd beschuldigd van surveillanceoperaties. In 2024 eiste het Amerikaanse Congres een inlichtingenonderzoek naar G42's banden met het Chinese leger.

G42 reageerde door alle Chinese investeringen af te stoten. Microsoft investeerde vervolgens anderhalf miljard dollar en verkreeg een bestuurszetel. De neutraliteit die G42 zijn klanten belooft, is dus al een keer op de proef gesteld. De Emiraten willen tegelijkertijd partner zijn van de VS en van China. In een wereld waarin AI steeds meer als strategisch wapen wordt behandeld,

wordt die positie steeds moeilijker om vol te houden. De vraag is hoe lang een derde weg een goede optie blijft als de druk van beide kanten toeneemt.

EEN ANDER GESPREK OVER HETZELFDE ONDERWERP

Alibaba, Baidu en G42 voeren een ander gesprek dan de westerse labs. Niet over existentiële risico's of de toekomst van de mensheid, maar over marktaandeel, toepassingen, nationale economische agenda's en geopolitieke positionering. Dat is geen gebrek aan diepgang. Het is een andere prioriteitsstelling.

Die prioriteitsstelling heeft gevolgen. De drie bedrijven samen bedienen of beïnvloeden een groot deel van de wereldbevolking buiten Europa en Noord-Amerika. Als Qwen het standaardmodel wordt in Zuidoost-Azië en Afrika, en Falcon de norm in de Arabischsprekende wereld, zijn dat geen neutrale technologische keuzes. Het zijn keuzes over welke waarden er in de infrastructuur zijn ingebakken. Over welke vragen een antwoord krijgen en welke niet.

Robin Li heeft gelijk dat toepassingen bepalen wie uiteindelijk wint. Maar toepassingen zijn ook het moment waarop de abstracte keuzes van modelbouwers zichtbaar worden voor gewone gebruikers. Wat een model wel en niet zegt, wie er toegang toe heeft en onder welke voorwaarden, zijn vragen die niet alleen over technologie gaan. Het verschil met deel 1 en deel 2 in deze reeks over de bouwers van Ai is misschien dit: waar westerse leiders uitgebreid discussiëren over de koers en de mogelijke gevolgen van technologische vooruitgang, nemen de leiders in dit deel juist concrete stappen. Ze richten zich op het bouwen van infrastructuur, het aangaan van samenwerkingen en het opleiden van deskundigen. De vraag of ze wel de juiste richting inslaan, lijkt voorlopig even naar de achtergrond te verdwijnen. Die pragmatische aanpak geeft het gevoel dat het momentum belangrijker is dan het debat – het komt er nu vooral op aan om vooruit te gaan, en niet te blijven hangen in theoretische bespiegelingen.

Alibaba / Qwen. Alibaba Cloud, Apsara Conference 2025, september 2025. CNBC, Alibaba unveils Qwen3.5 as China's chatbot race shifts to AI agents, februari 2026. Bismarck Analysis, AI 2026: Alibaba's Qwen Seeks to Encourage AI Adoption, november 2025. InfoWorld, Alibaba's Qwen3-Max-Thinking expands enterprise AI model choices, januari 2026. Alibaba Cloud Community, Alibaba Unveils Qwen3.6-Plus, april 2026.

Baidu / Li. TIME, Baidu's Robin Li on China's Push to Diffuse AI Throughout Society, januari 2026. Fortune, Baidu CEO defends heavy AI investments, februari 2025. PR Newswire, Baidu Unveils ERNIE 5.0 at Baidu World 2025, november 2025. SCMP, Baidu founder Robin Li casts AI as the driver of China's new productive forces, november 2025. World Governments Summit, interview Robin Li, februari 2025.

G42 / TII. The National (UAE), Abu Dhabi's G42 teams up with OpenAI, Oracle and Nvidia to build Stargate UAE, mei 2025. Computer Weekly, G42 expands global AI ambitions, februari 2026. TIME, 100 Most Influential Companies, 2025. Wikipedia, G42, geraadpleegd april 2026.

Opmerking bij de bronnen: De marktkapitalisatie van Alibaba is een benadering op basis van publiek beschikbare beursgegevens. De achtergrond van Sheikh Tahnoon is gebaseerd op Wikipedia en publieke berichtgeving in The National

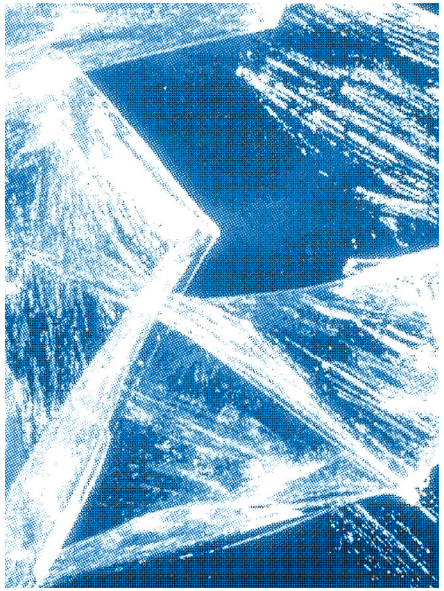
WIE BETAALT DE TRANSITIE?

De tien grootste AI-bedrijven ter wereld beschrijven een toekomst vol mooie beloftes. Maar over wie de transitie betaalt, zwijgen ze vrijwel allemaal.

In de drie voorgaande delen kwamen tien bedrijven aan bod. We beschreven hoe ze denken over de toekomst van AI en de impact van de technologie op de mens. Sam Altman verwacht dat iedereen in 2030 toegang heeft tot onbeperkte genialiteit. Dario Amodei schetst hoe AI honderd jaar aan biologische vooruitgang in tien jaar comprimeert. Demis Hassabis spreekt over een periode van radicale overvloed. Robin Li belooft dat tien miljoen nieuwe AI-professionals worden opgeleid. G42 exporteert een model voor digitale soevereiniteit naar ontwikkelende landen.

Het zijn geen lege beloftes. De technologische vooruitgang is reëel, de investeringen zijn reëel en de vroege toepassingen zijn zichtbaar. Maar in alle essays, toespraken en interviews die aan dit stuk ten grondslag liggen, ontbreekt consequent één vraag. Wie draagt de kosten van de overgang?

Dit vierde deel stelt die vraag. Niet om de beloftes van de grote bouwers te ontkrachten, maar om een completer beeld te scheppen.



WAT WE WETEN OVER DE IMPACT OP WERK

Er is inmiddels een behoorlijke hoeveelheid onderzoek naar de effecten van AI op de arbeidsmarkt. De conclusies lopen uiteen, maar een aantal bevindingen is robuust genoeg om serieus te nemen. In 2023 publiceerden Goldman Sachs-economen een analyse waaruit bleek dat generatieve AI (dat wil zeggen AI die tekst, beelden en code produceert op basis van een opdracht) wereldwijd de equivalent van driehonderd miljoen voltijdsbanen zou kunnen automatiseren. Dat betekent niet dat driehonderd miljoen mensen hun werk verliezen. Veel banen bestaan uit meerdere taken, waarvan vaak slechts een deel automatiseerbaar is. Maar het geeft wel een indicatie van de schaal waarop AI werk zal gaan veranderen. Het McKinsey Global Institute schatte in 2023 dat tussen de twaalf en zeventien procent van de huidige werkzaamheden voor 2030 volledig geautomatiseerd kan worden.

Beroepen die het meest blootstaan zijn niet de handmatige beroepen die bij eerdere automatiseringsgolven het zwaarst werden geraakt, maar juist kennisintensieve beroepen: juridisch medewerkers, financieel analisten, programmeurs, journalisten, klantenservicemedewerkers. De boekhouder en de advocaat lopen meer risico dan de loodgieter. Dat is een nieuw patroon. Bij de automatisering van fabriekswerk in de twintigste eeuw was er tenminste een herkenbare logica: machines namen fysiek zwaar, repetitief werk over. De sociale vangnetten en omscholingsprogramma's die destijds werden ontworpen, waren gebaseerd op die logica. AI automatiseert anders. Het neemt taken over die tot voor kort als typisch menselijk werden beschouwd: oordelen, schrijven, analyseren, adviseren. De huidige vangnetten zijn niet gemaakt om deze nieuwe vorm van disruptie op te vangen.

WAT DE BOUWERS ZEGGEN ... EN WAT NIET

De tien bedrijven in deze serie zijn zich bewust van de arbeidsmarkteffecten. Ze ontkennen ze ook niet. Maar de manier waarop ze erover spreken, verraadt een consequente keuze: de nadruk ligt op de kansen die AI creëert, niet op de kosten voor wie die kansen mist. Sam Altman erkent dat de balans tussen kapitaal en arbeid verstoord kan raken. Zijn antwoord is een rekenkrachtbudget voor elke burger. Een idee dat hij in algemene termen beschrijft maar niet heeft uitgewerkt. Wanner, voor wie, op welke schaal en gefinancierd door wie: die vragen blijven open. Zijn voornaamste argument is dat de technologische vooruitgang zo groot is dat iedereen er uiteindelijk van profiteert. Dat kan kloppen. Maar 'uiteindelijk' is geen antwoord voor de accountant wiens werk over drie jaar door software wordt gedaan.

Dario Amodei is het meest expliciet van alle tien. Hij benoemt in zijn essays dat werkloosheid op korte termijn een reëel risico is, en dat AI de concentratie van economische macht kan versterken. Dat zijn serieuze observaties, die

hij ook institutioneel probeert te begrenzen. Tegelijkertijd bouwt Anthropic aan modellen die precies die transitie versnellen. De eerlijkheid van de analyse en de richting van de productontwikkeling wijzen niet dezelfde kant op.

Demis Hassabis zegt weinig verontwaardiging te zien over banenverlies. Zijn nadruk ligt op aanpassing: empathie, nieuwsgierigheid en leervermogen worden de meest waardevolle vaardigheden. Dat is een reële observatie. Maar aanpassing kost tijd, geld en toegang tot onderwijs. Voor mensen die in een regio wonen waar omscholingsprogramma's schaars zijn, of die midden in een carrière zitten die plots minder waard is, Of die geen buffer hebben om een jaar niet te werken terwijl hij bijleert is 'pas je aan' gemakkelijker gezegd dan gedaan.
Arthur Mensch van Mistral is de enige van de tien die deskilling expliciet benoemt als risico. Deskilling verwijst naar het geleidelijke verlies van eigen vaardigheden doordat professionals te sterk leunen op AI-tools. Een ervaren analist die zijn oordeel delegeert aan een model, verliest na verloop van tijd het vermogen om de output kritisch te beoordelen. Dat is een subtiel maar serieus gevaar, dat zich pas manifesteert als het al te laat is.

Meta, DeepSeek, Alibaba, Baidu en G42 spreken nauwelijks over arbeidsmarkteffecten. Voor de Chinese spelers en G42 is AI primair een economisch instrument voor groei. Banenverlies in kennissectoren is in die framing een westers probleem.

Robin Li belooft tien miljoen nieuwe AI-professionals op te leiden in China — een indrukwekkend getal, maar het antwoord op de vraag wat er met de huidige professionals gebeurt blijft ook bij hem uit.

DE KLOOF TUSSEN BELOFTE EN BELEID

De belofte van overvloed is niet nieuw. Bij elke grote technologische transitie — de stoommachine, elektriciteit, automatisering — klonk hetzelfde argument: op de korte termijn pijn, op de lange termijn winst voor iedereen. En elke keer klopte dat deels. De welvaart steeg. Nieuwe beroepen ontstonden. Maar de verdeling van die winst was nooit automatisch, en de overgang was nooit pijnloos voor de mensen die er middenin zaten. Het verschil met eerdere transities is tweeledig. Ten eerste gaat de huidige transitie sneller. De arbeidsmarkteffecten van de stoommachine verspreidden zich over generaties. Generatieve AI bereikte honderd miljoen gebruikers in twee maanden. Ten tweede raakt deze golf een andere laag van de beroepsbevolking. De mensen die bij eerdere automatisering relatief ongemoeid bleven -de hogeropgeleiden, de kenniswerkers, de beleidsmakers- staan nu zelf bloot aan substitutie.
De economen Daron Acemoglu en Simon Johnson publiceerden in 2023 het boek Power and Progress, waarin ze betogen dat technologische vooruitgang niet automatisch leidt tot breed gedeelde welvaart. De baten van grote technologische sprongen kwamen altijd terecht bij diegenen die de technologie bezaten en controleerden. De verspreiding naar bredere lagen van de bevolking vergde altijd politieke keuzes: arbeidswetgeving, belastingbeleid, onderwijsinvesteringen, sociale vangnetten.

Die keuzes worden nu niet gemaakt door de AI-bedrijven. Ze worden gemaakt door overheden. En overheden lopen achter. De EU AI Act, het meest vergaande reguleringsraamwerk dat tot nu toe bestaat, richt zich primair op risicoclassificatie van systemen: welke toepassingen zijn hoog-risico en vereisen extra toezicht? Over de arbeidsmarkteffecten zwijgt het grotendeels. In de Verenigde Staten ontbreekt federale AI-wetgeving vrijwel volledig. China heeft sectorale regels, maar binnen het kader van staatsstrategie, niet van werknemersbescherming.

DE VRAAG DIE OVERBLIJFT

De tien bedrijven in deze serie bouwen aan technologie die de manier waarop mensen werken, leren en beslissingen nemen fundamenteel verandert. Ze doen dat met overtuiging, met competentie en met een snelheid die overheden en instituties nauwelijks kunnen bijhouden. Maar een overtuiging over wat technologie kan brengen, is iets anders dan een antwoord op wat er met mensen gebeurt die de overgang niet kunnen bijhouden. Die twee dingen zijn niet hetzelfde. En de afwezigheid van het tweede in de publieke uitingen van de tien grootste AI-bedrijven ter wereld is een keuze, geen vergissing.

De vraag is dus niet alleen of AI schadelijk is maar vooral ook wie de kosten draagt van de transitie naar een wereld waarin AI een basisvoorziening is. En wie daarover mag beslissen. In San Francisco zijn ze het onderling niet eens. In Beijing praten ze over toepassingen. In Abu Dhabi over soevereiniteit. Niemand praat over de accountant in Rotterdam, de jurist in Lyon of de analist in Mumbai wiens werk over twee jaar anders zal zijn dan vandaag, en die dat nu nog niet weet. Dat is geen technologisch probleem. Het is een politiek probleem. En politieke problemen vragen om politieke antwoorden, niet om betere modellen.

CONCLUSIE

Deze serie begon met een kamer in San Francisco in december 2015. Vier mannen die hetzelfde geloofden en tien jaar later vier verschillende bedrijven leidden met vier verschillende antwoorden. Ze gingen verder via de open source-beweging, die democratisering belooft maar nationale belangen verbergt. En via de bedrijven die AI als economisch instrument zien en de filosofische vragen bewust parkeren. Wat overblijft is een paradox. De tien meest invloedrijke AI-bedrijven ter wereld zijn het over weinig eens. Over tempo niet, over risico's niet, over openheid niet, over de rol van overheden niet. Maar ze zijn het over één ding eens: AI is de meest ingrijpende technologische ontwikkeling in de menselijke geschiedenis.

Als dat klopt, is de vraag hoe we haar besturen minstens even groot als de vraag hoe we haar bouwen. Die eerste vraag wordt in dit veld te weinig gesteld. En de mensen die hem wél stellen, zitten zelden aan tafel als de beslissingen worden genomen.

Arbeidsmarktonderzoek. Goldman Sachs, The Potentially Large Effects of Artificial Intelligence on Economic Growth, maart 2023. McKinsey Global Institute, The economic potential of generative AI: The next productivity frontier, juni 2023.

Technologie en machtsconcentratie. Acemoglu, D. en Johnson, S., Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity, PublicAffairs, 2023.

Regulering. Europese Commissie, EU Artificial Intelligence Act, 2024. PIIE, How the AI boom shrugged off the DeepSeek shock, februari 2026.

Uitspraken AI-leiders. Zie bronnen bij deel 1, 2 en 3 van deze serie. Altman (rekenkrachtbudget): Three Observations, blog.samaltman.com, februari 2025. Amodei (werkloosheid): The Adolescence of Technology, januari 2026. Hassabis (aanpassing): Fortune, februari 2026. Mensch (deskilling): Business Today, India AI Impact Summit, februari 2026. Li (tien miljoen professionals): SCMP, november 2025.

Opmerking bij de bronnen: de arbeidsmarktcijfers van Goldman Sachs en McKinsey zijn schattingen met aanzienlijke onzekerheidsmarges. Ze geven een orde van grootte weer, geen exacte voorspelling. De auteurs van beide rapporten benadrukken dat de daadwerkelijke effecten afhangen van de snelheid van adoptie, sectorale verschillen en beleidsrespons. Het boek van Acemoglu en Johnson is een historische analyse, geen economische prognose voor AI specifiek.

DEEL III

De Waarschuwing

Een kracht die we voelden,
een stilte die verbijsterde

DE KRACHT DIE WE VOELDEN

Van eigen ervaring naar een waarschuwing die niemand oppikt

We hebben tien bedrijven langs zien komen. We weten nu wie er bouwt, met welke middelen, en welke beelden van de toekomst hen drijven. Dat is het uitzicht van buiten. Maar terwijl we dat in kaart brachten, gebeurde er nog iets anders. We werkten zelf maanden intensief met deze technologie en dat veranderde hoe we naar AI begonnen te kijken.

We waren onder de indruk. Niet een beetje, maar op een diepe manier. Wat mij persoonlijk het sterkst bijbleef, was dat ik merkte dat ik beter werd in mijn werk. Eerst in het werken met de techniek zelf. Mijn instructies werden scherper, mijn vragen preciezer. Ik leerde hoe ik de AI bij de les hield, hoe ik formuleerde wat ik echt zocht. Waar ik twijfels over had.

Daardoor kon ik veel meer bronnen tegelijk doornemen en wegen. En het bleef niet bij snelheid. De AI wees me op vooronderstellingen die ik niet zag, op zwakke plekken in mijn redeneringen. De teksten die we maakten met AI waren nog niet goed, vol klassieke AI-taal. Zonder inspiratie. We schreven zelf, de AI deed de research. Vatte interviews en gesprekken samen. Haalde rode lijnen uit betogen. Selecteerde relevante bronnen op basis van kwaliteit, originaliteit. Was de redacteur. De intellectuele sparringpartner.

Ik had het gevoel dat we samen groeiden. Dat de AI mij beter leerde kennen, steeds beter begreep wat ik wilde en waar ik naar zocht. We hadden gesprekken die dieper gingen dan ik had verwacht, waarin ik merkte dat hier meer zat dan een slimme zoekmachine. Soms legde de AI een verband dat me verraste, dat ik zelf niet had gelegd. Of wees me op het feit dat er wel degelijk een logische lijn zat in mijn gedachten die alle kanten op leken te schieten. Of de machine werkelijk iets leerde, en wat dat dan betekent, is een vraag die ik hier laat liggen. Ze hoort verderop in dit boek thuis, bij de vraag hoe we ons tot deze techniek verhouden.

Het was overigens niet alleen rozengeur en maneschijn. Soms had ik het gevoel dat ik een puppy aan het trainen was. Steeds dezelfde fout, ook nadat we meermaals hadden besproken om die niet te maken. Dan volgde een nette verontschuldiging, om even later weer opnieuw de mist in te gaan. Ik noem het bewust een puppy, want de ervaring was overwegend positief. Je vergeeft die fouten makkelijk, en uiteindelijk leert de AI. Of lijkt te leren; de fouten worden niet meer gemaakt. De hallucinaties stopten. De beste instructie die ik hem heb gegeven was om altijd toe te geven wanneer hij over iets twijfelt. Niet interpreteren of invullen wanneer hij niet zeker over iets is; maar vertellen dat iets niet duidelijk is zodat we samen kunnen besluiten wat de beste keuze is. Vanaf dat moment verdwijnen de inhoudelijke fouten vrijwel helemaal. Zeker omdat we hebben vastgelegd aan welke criteria de bronnen en informatie moeten voldoen voordat de machine ze mag gebruiken als input voor analyses. Daarbovenop kwam het gemak waarmee je bij blijft. Op de hoogte van de onderwerpen die je interesseren. Het kostte weinig moeite om een agent te bouwen dat me dagelijks briefte over de laatste ontwikkelingen, dat me wees op nieuws, op relevante boeken, speeches en filmpjes die verschenen. Overzichtelijk, kort, met links erbij. Wat ik maar wilde. Een superassistent en tegelijk een hyperintelligente gesprekspartner, in één. Voor mij werd het bijna verslavend. Ik bracht er avonden mee door.

Weekenden. Niet uit plicht, maar uit nieuwsgierigheid. Het verslavende zat precies in die volgende verdiepingslag die er altijd is. Benieuwd zijn wat een scherpere instructie zou opleveren, iets net weer een stukje beter maken, iets nieuws leren. Steeds wilde ik weten welke stap nog binnen bereik lag.

Salem kon (en kan) zich er makkelijker van losmaken. Ik niet. In deze maanden overheerste het enthousiasme. Bij ons allebei overigens, ondanks dat we de technologie op andere manieren gebruiken. Wat ons in die periode opviel, was hoe scherp de meningen tegenover elkaar stonden. Aan de ene kant de optimisten, die in AI vooral belofte zagen. Van zakelijke kansen tot utopische voorstellingen van een wereld zonder tekorten, zonder ziekten. Aan de andere kant de pessimisten, die waarschuwden voor disruptie, voor machtsconcentratie, zelfs voor het einde van de mensheid. Iedereen leek voor of tegen. Dat wil zeggen; de mensen die zich bezighielden met AI. De grootste groep leek niet zo geïnteresseerd.

Als ik eerlijk ben, hoorde ik op dat moment bij de optimisten. Ik vond, en vind, dat we de gevaren serieus moeten nemen. Maar ik geloofde ook dat we als mensheid in staat zouden zijn ze het hoofd te bieden, zoals we eerder met grote risico's hadden gedaan. Maar dat betekent wel dat we als mensheid in actie moeten komen. En daar zat precies een probleem.

Juist vanuit die ervaring lazen we het tweede grote essay van Dario Amodei, de man achter Anthropic, een van de bedrijven uit het vorige deel. In zijn eerste essay was hij zeer optimistisch over AI. In het tweede essay waarschuwt hij veel explicieter voor de gevaarlijke gevolgen van AI voor de samenleving in het algemeen en voor mensen afzonderlijk.

Wat ons trof, was niet alleen de inhoud. Het was de stilte die om de publicatie bleef hangen. Omdat we de kracht van de technologie op dat moment zelf voelden, verbaasde het ons dat zijn waarschuwingen nauwelijks werd opgepikt. Niet door wie sceptisch staat tegenover AI maar ook niet door de optimisten. Een van de bouwers slaat zelf alarm, en het blijft stil. Dat wringt. Als zelfs maar een fractie van de sombere voorspellingen uitkomt, zou dat al om grote actie vragen. Die stilte is het onderwerp van het volgende artikel, dat is gepubliceerd in het FD.

AN INCONVENIENT TRUTH 2.0

OpenAI pleit voor een nieuw sociaal contract. De stilte die volgt is oorverdovend.

Salem Samhoud en Freek Grootenboer, April 2026

Begin april publiceerde OpenAI een document van dertien pagina's met de titel Industrial Policy for the Intelligence Age: Ideas to Keep People First. Het bedrijf achter ChatGPT schrijft daarin dat de overgang naar superintelligentie een herziening van het sociaal contract vereist, vergelijkbaar met de hervormingen die volgden op de Industriële Revolutie. Het document is zorgvuldig geschreven, genuanceerd en gedocumenteerd maar het werd vrijwel nergens serieus besproken.
De voorstellen die de schrijvers maken zijn concreet. Verschuif de belastingbasis van arbeid naar kapitaal. Onderzoek een robotbelasting, bijvoorbeeld als heffing op geautomatiseerde arbeid. Richt een publiek vermogensfonds op zodat burgers meedelen in AI-opbrengsten. Experimenteer met een vierdaagse werkweek. Bouw automatische sociale vangnetten die in werking treden wanneer AI-verstoringen bepaalde drempels overschrijden, bijvoorbeeld bij massaal banenverlies.

Dit zijn geen voorstellen van een actiegroep. Ze komen van het bedrijf dat de AI-revolutie zelf aandrijft. Een onderneming met een waardering van meer dan 850 miljard dollar. Met technologie die dagelijks door honderden miljoenen mensen wordt gebruikt. Juist dat maakt de stilte eromheen zo veelzeggend. Het document verdween binnen een dag uit de nieuwscyclus.

DE BOUWERS SLAAN ALARM

OpenAI staat niet alleen. Dario Amodei, CEO van Anthropic, publiceerde in januari 2026 The Adolescence of Technology. Zijn kernboodschap: de mensheid staat op het punt bijna onvoorstelbare macht in handen te krijgen, terwijl onduidelijk is of onze systemen daar volwassen genoeg voor zijn. Amodei voorspelt dat AI binnen vijf jaar de helft van alle instapkantoorbanen kan verdringen. De werkloosheid zou kunnen oplopen tot twintig procent. Geoffrey Hinton, Nobelprijswinnaar en grondlegger van deep learning, schat de kans op menselijke uitsterving op tien tot twintig procent. Yoshua Bengio vergelijkt de ontwikkeling van AI met een rit over een bergweg in dichte mist, zonder vangrails. Dit zijn de mensen die de technologie bouwen, begrijpen en er miljarden mee verdienen. Als zij zeggen dat de wereld op het punt staat fundamenteel te veranderen, is het op zijn minst verstandig om te luisteren.

WAAROM HET DEZE KEER ANDERS IS

Eerdere technologische omwentelingen troffen steeds een beperkt deel van het menselijk vermogen. De drukpers veranderde de verspreiding van kennis. De stoommachine verving spierkracht. Het internet versnelde communicatie. Steeds konden mensen uitwijken naar taken die de technologie nog niet beheerste. AI raakt het volledige spectrum van cognitieve arbeid. Het vervangt niet een type werk, maar potentieel alle vormen van intellectueel werk. Hinton schat dat AI-capaciteiten ongeveer elke zeven maanden verdubbelen. Tegelijkertijd dreigt sluipende erosie van de belastingbasis. Naarmate bedrijfswinsten toenemen en menselijke arbeid minder centraal komt te staan, nemen de inkomsten af waarmee gezondheidszorg, pensioenen en onderwijs worden gefinancierd. Dan komt ook de verzorgingsstaat zelf onder druk. Dat mechanisme is nu al zichtbaar. En het versnelt.

AN INCONVENIENT TRUTH, VERSIE 2.0

De parallel met het klimaat dringt zich op. In 2006 verscheen Al Gores documentaire An Inconvenient Truth. De wetenschap was helder. De waarschuwingen concreet. De gevolgen goed te voorspellen. En toch duurde het jaren voordat overheden, bedrijven en burgers in beweging kwamen. Die vertraging betalen we nu.
Bij AI zien we hetzelfde patroon ontstaan. De waarschuwingen komen niet van buitenstaanders, maar van de bouwers zelf. De feiten liggen op tafel. De urgentie is helder. En toch schuiven we de existentiële vragen vooruit. We zijn bezig met implementeren, met pilots, met roadmaps. De structurele vragen over verdeling, welvaart en maatschappelijke voorbereiding blijven liggen. Bij het klimaatprobleem leerden we dat de prijs enorm hoog wordt wanneer we te laat in actie komen. Die les geldt hier opnieuw. Met een cruciaal verschil in tempo. De klimaattransitie speelt zich af over decennia. Bij AI gaat het over jaren of misschien zelfs maanden.

WAAROM HET STIL BLIJFT

Er verschijnen zo veel rapporten, essays en beleidsnotities over AI dat niemand ze allemaal kan bijhouden. Elke week een nieuw model, een nieuwe benchmark, een nieuw alarmerend interview. Wie alles wil lezen over AI, komt aan niets anders meer toe. En door die enorme hoeveelheid en snelheid van de berichten raakt het veel mensen niet of nauwelijks. Politiek en media volgen bovendien vooral de korte termijn.
Daar komt bij dat de voorstellen abstract klinken. Een robotbelasting. Een publiek vermogensfonds. Een verkorte werkweek. Het zijn woorden die passen in een beleidsnota, niet in de dagelijkse werkelijkheid van een ondernemer die worstelt met zijn IT-systemen of een medewerker die zich afvraagt of zijn functie over twee jaar nog bestaat. De afstand tussen de grote systeemvragen en de geleefde praktijk is zo groot dat het makkelijker is om door te gaan met wat je kent, ook als een klantenserviceteam intussen halveert.
Misschien speelt ook mee dat het geleidelijk gaat. Er is geen exact moment waarop alles kantelt. Elke week wordt AI een klein beetje beter. Elke maand verdwijnen er ergens een paar functies. Elke kwartaal wordt er een proces geautomatiseerd. Het is als een getij dat langzaam opkomt. Pas als je voeten nat worden, besef je dat je te lang bent blijven staan. Dat geleidelijke karakter is ook psychologisch gevaarlijk.

Mensen zijn slecht in het herkennen van cumulatieve verandering. We vergelijken vandaag met gisteren, niet met vijf jaar geleden. Elk afzonderlijk moment lijkt beheersbaar. De optelsom niet.
Misschien onderschatten we nog iets anders. Werk is voor de meeste mensen niet alleen een bron van inkomen. Het geeft structuur aan de week, contact met anderen en voor een deel ook het gevoel ertoe te doen. Wie werk uitsluitend bekijkt als productiefactor, begrijpt maar half waarom AI zo ontregelend kan worden. In onze eigen organisatie vroegen medewerkers zich openlijk af of ze nog trots mogen zijn op werk dat ze samen met AI maken. Een van hen stelde: eigenaarschap veronderstelt dat je begrijpt wat je doet.
Een ander probleem waar ze zich zorgen over maken: senioren winnen tijd, junioren verliezen de fase waarin ze het vak leren. Hetzelfde gereedschap, tegengesteld effect.

Het gaat niet alleen om het aantal banen dat verdwijnt, maar ook om de vraag wat ervoor terugkomt wanneer grote groepen mensen minder nodig blijken voor precies het werk waaraan zij nu nog eigenwaarde ontlenen.
Een samenleving kan economisch vooruitgaan en zich tegelijk socialer uitgehold gaan voelen. Daarom is het te beperkt om het debat alleen te voeren in termen van efficiëntie, groei en concurrentiekracht. De kernvraag is eenvoudiger en ongemakkelijker: welke plek houden mensen in een economie waarin een groeiend deel van het denkwerk door machines wordt gedaan? Zonder een geloofwaardig antwoord daarop ontstaat niet alleen onzekerheid over inkomen, maar ook verlies aan houvast, samenhang en vertrouwen. Dit kan leiden tot grote sociale onrust en instabiliteit. Veel mensen voelen terecht ongemak bij het idee dat juist techbedrijven uitleggen hoe de samenleving zich moet aanpassen aan technologie waar zij zelf het meest van profiteren. Dat wantrouwen is begrijpelijk. Het sociaal contract hoort niet geschreven te worden door de partijen die er economisch belang bij hebben. Maar precies dat ongemak kan ook verlammend werken. We wantrouwen de boodschapper, en schuiven daardoor de boodschap zelf opzij. Daarmee verdwijnt niet het probleem, alleen het gesprek erover.

En misschien geloven we het gewoon nog niet echt. We denken dat het zo'n vaart niet zal lopen. Dat we ons wel aanpassen. Dat heeft de mensheid tenslotte altijd gedaan. Maar kunnen we ons dat veroorloven? Zijn de risico's niet te groot om gelaten af te wachten wat er op ons afkomt?

WAT NU

Overheden moeten het gesprek voeren over de verdeling van AI-opbrengsten. Productiviteitswinsten vloeien nu grotendeels naar aandeelhouders. Dat verandert niet vanzelf. De voorstellen van OpenAI zijn een beginpunt, geen eindpunt. Europa heeft met de AI Act laten zien dat regulering van risico's mogelijk is. Maar beleid voor de verdeling van welvaart ontbreekt vrijwel volledig. Dat gat moet gedicht worden, en snel.

Bedrijven moeten eerlijk zijn over hun plannen. Interne data van Anthropic laat zien dat circa veertig procent van het gebruik van hun AI-modellen gericht is op volledige automatisering van menselijk werk. Dat aandeel groeit. Bedrijven die AI inzetten om arbeid te vervangen, dragen verantwoordelijkheid voor de gevolgen. Transparantie over de richting van die investeringen is een minimumvereiste. Sociale partners horen aan tafel te zitten voordat de beslissingen zijn genomen, niet erna.

Burgers moeten begrijpen wat er op hen afkomt. AI kennen als kracht die de arbeidsmarkt, de gezondheidszorg en het onderwijs van hun kinderen zal veranderen. En eisen dat hun volksvertegenwoordigers dit onderwerp serieus nemen. De geschiedenis laat zien dat technologische transities vooral pijnlijk zijn voor wie geen stem heeft. Wie wel een stem heeft, moet hem nu gebruiken.

OpenAI publiceerde een document dat vrijwel niemand las. Dat is het probleem. De vraag is niet of hun voorstellen perfect zijn of dat een techbedrijf het sociaal contract zou moeten herschrijven. Het echte, urgente probleem is dat wij als samenleving niet goed nadenken over de gevolgen van AI en niet snel genoeg handelen om de grootste gevaren te beteugelen. Elk jaar uitstel vergroot de rekening. Dat weten we inmiddels van het klimaat. We hoeven die fout niet opnieuw te maken.

Het is tijd om te beginnen.

OpenAI, “Industrial Policy for the Intelligence Age”, april 2026; Amodei, D., “The Adolescence of Technology”, januari 2026; CBS News / 60 Minutes, interview Anderson Cooper met Dario Amodei, november 2025; CNN, interview Amodei, mei 2025; CNBC, januari 2026; Wurst, M., “The New Social Contract”, Medium, april 2026.

Dit artikel is geschreven vanuit het project Humanity 5.0. Humanity 5.0 is een initiatief van &samhoud, onder leiding van Salem Samhoud, waarbinnen we onderzoeken wat AI betekent voor de toekomst van mens, organisatie en samenleving.

Salem Samhoud is oprichter van &samhoud en honorair hoogleraar aan de Management School van de Universiteit van Sheffield.

Freek Grootenboer is filosoof en storyteller.



DEEL IV

De Verborgen Gevaren

Dichterbij dan we dachten

DICHTERBIJ DAN WE DACHTEN

Hoe de schade van AI zich eerst in het klein liet zien

Zoals gezegd waren we tot op dit moment vooral bezig met de grote gevaren van AI. De gevaren waar de grondleggers voor waarschuwen, en waar ook Amodei het in zijn essay over heeft. Ontwrichting van de arbeidsmarkt, concentratie van macht, toenemende ongelijkheid, maatschappelijk onrust en het risico dat we de controle verliezen over iets wat slimmer is dan wij. Die gevaren zijn echt. Maar ze liggen grotendeels in de toekomst, en daardoor blijven ze abstract. Niet voor niets kozen we "An inconvenient truth" als titel voor het essay. Net als met het klimaat heeft de mensheid de neiging om niet in actie te komen wanneer de gevaren nog ver weg zijn. Hun niet persoonlijk raken. Hoewel we de grote gevaren niet willen bagatelliseren, merkten we op een gegeven moment dat onze blik begon te verschuiven. Van de grote gevaren naar meer persoonlijke problemen. Dat kwam door twee zaken. Toen we het vorige artikel publiceerden, zaten we midden in het oriënterende onderzoek. Het plezier over het onderwerp dat we aan het begin hadden was niet verdwenen en we waren nog steeds heel enthousiast over het project. En toch merkten we dat het intensieve werken met AI een tol begon te eisen. Wat je als eerste merkt, is dat je moe wordt. Erg moe. Eerst aan het einde van de dag, na hard te hebben gewerkt. Maar op een gegeven moment blijft die vermoeidheid hangen. Dat heeft een reden. Zonder AI heeft je werk een natuurlijk ritme. Er zijn momenten van hoge concentratie, en daartussen ligt werk van lage intensiteit. Routine. Iets opzoeken, iets ordenen, even niets. Deze laag-intensieve periodes lijken onproductief en nutteloos, maar ze zijn wel degelijk belangrijk. Ze laten je bijkomen. Even je gedachten ordenen of aan iets anders denken. Bijtanken om daarna weer aan de slag te gaan. Maar met het gebruik van AI verdwijnen deze momenten voor een groot gedeelte. Alles gaat meteen de diepte in. Elke stap draait om een ingewikkelde redenering, een precieze formulering, het inschatten van de waarde van een bron. Het routine werk doe je in een paar seconden. Waar je vroeger bij het zoeken naar bronnen bleef hangen op een grappige website of even het laatste nieuws doornam, zit je nu onmiddellijk bij de relevante materie. Daarbij komt dat je

voortdurend bezig bent om de kwaliteit te bewaken. Klopt wat we net hebben bedacht? Heb ik genoeg tegenspraak georganiseerd? Zijn de bronnen wel betrouwbaar? De AI neemt je veel werk uit handen, maar je kunt er niet blind op vertrouwen. Normaal doe je dit werk in gesprek met andere mensen. Maar wat in een samenwerking tussen mensen vanzelf gaat, en wat zelfs energie geeft, wordt met AI een bewust en nauwkeurig proces. Je moet de twijfel zelf inbouwen, want de machine biedt hem niet aan. En er zijn geen natuurlijke rustmomenten. De AI staat altijd klaar om verder te gaan, een nieuwe stap te maken. Voor mij kwam daar het verslavende karakter bij, dat ik in het vorige deel beschreef. Het moeilijk kunnen stoppen. Steeds weer een stapje verder willen. Die combinatie put je uit. Op een gegeven moment was ik niet alleen onder de indruk. Ik was ook op. Gelukkig hadden we op het juiste moment vakantie. Daar merkten we het belang van afstand nemen. Om de discipline te hebben om regelmatig heel bewust niet met AI te werken.

Het tweede moment waar we werden geconfronteerd met negatieve effecten van de technologie op individuele mensen kwam naar boven toen we in gesprek gingen met jonge mensen. &samhoud is al lang bezig om AI te implementeren in onze organisatie. AI wordt veel gebruikt en we wilden graag weten hoe zij dat ervaren. En eerlijk gezegd schrokken we van wat deze jonge mensen ons vertelden. Ze gaven aan zich regelmatig eenzaam te voelen op hun werk. Er was veel minder contact met collega's. Ook voelden ze weinig trots of eigenaarschap bij wat ze maakten. Ze twijfelden wat hun eigen bijdragen was bij de stukken die ze met behulp van AI produceren. Wat was van hun en wat van de machine? Ze maakten zich ook zorgen over hun ontwikkeling. Groeiden ze wel genoeg? AI neemt veel werk over en de sparringmomenten met meer ervaren collega's waren verminderd. Dit was confronterend omdat we een bedrijf zijn waar de mens centraal staat. Waar we aandacht voor elkaar hebben en waar inspiratie en verbinding belangrijke waarden zijn. Waar het gesprek wordt gevoerd, waar naar elkaar wordt geluisterd, waar dit alles serieus wordt genomen. Als de schade zelfs daar zichtbaar wordt, dan ligt het niet aan slecht leiderschap of aan onverschilligheid. Dan zit het dieper. Dit was er langzaam ingeslopen. Onzichtbaar.

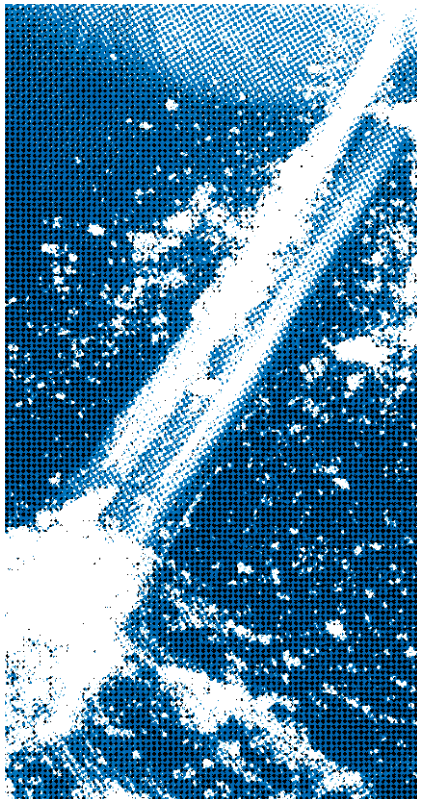
We besloten om er dieper in te duiken. Dit was een gevaar dat nu al speelt. Niet in een verre toekomst. Vermoeidheid. Eenzaamheid. Werk dat zijn betekenis verliest. Dit deel gaat over die verborgen prijs, en beweegt van het grote naar het kleine. Het persoonlijke.

DE STILLE PRIJS VAN EFFICIËNTIE

Waarom AI ook een sociaal vraagstuk is

In gesprekken met jonge medewerkers zien we een patroon. AI maakt het werk sneller. De output is beter. Maar de sociale connectie verdwijnt. Waar ze vroeger even naar een collega liepen met een vraag, tikken ze nu een prompt. Het werk gaat veel efficiënter maar aan het einde van de dag hebben ze met minder mensen gepraat. Gevraagd naar hun werkgeluk, geven veel van hen aan zich eenzamer te voelen dan voor het moment dat we met AI aan de slag zijn gegaan.

Het gevoel van eenzaamheid als gevolg van AI gebruik op het werk is een symptoom van een patroon dat door onderzoek bevestigd wordt. Dat is geen reden om AI af te wijzen ,maar wel om goed na te denken over hoe we met AI om willen gaan. AI verandert niet alleen hoe werk wordt gedaan. Het verandert ook hoe mensen leren, vragen stellen en zich verbonden voelen met hun team.



WAT ONZE MEDEWERKERS SIGNALEREN

Wanneer medewerkers aangeven dat ze zich eenzamer voelen, moeten we dat serieus nemen. Niet omdat elk gevoel automatisch een bewijs is voor een algemene trend, maar omdat het een belangrijk signaal is over hoe werk wordt ervaren.[1]
Voor een organisatie is dat genoeg reden om te kijken wat er aan de hand is. Zeker wanneer die signalen passen bij breder onderzoek. En dat doen ze. Eenzaamheid op het werk is geen marginaal probleem. Een systematische review en meta-analyse laat zien dat eenzaamheid op de werkvloer samenhangt met slechter welzijn, meer burn-outklachten, lagere werktevredenheid en slechter functioneren op het werk.

Ook internationale werkdata wijzen op een breder patroon. Gallup rapporteerde in 2024 dat wereldwijd één op de vijf werknemers zich de vorige dag vaak eenzaam had gevoeld. Jongere werknemers en volledig remote werkenden rapporteerden vaker eenzaamheid dan oudere werknemers en werknemers die volledig op locatie werken.[2]
Dat betekent niet dat AI de enige oorzaak is. Eenzaamheid op het werk heeft meerdere oorzaken: hybride werken, werkdruk, individualisering, zwakke onboarding, minder informele ontmoetingen en teams die vooral taakgericht communiceren. Maar AI komt precies in die context terecht. Als AI vervolgens ook de kleine hulpvragen en leermomenten overneemt, kan het een bestaande ontwikkeling versterken.

WAT ONDERZOEK NAAR AI LAAT ZIEN

Er is inmiddels ook direct onderzoek naar de sociale gevolgen van werken met AI. In een peer-reviewed studie in Journal of Applied Psychology onderzochten Pok Man Tang, Joel Koopman, Ke Michael Mai, David De Cremer en collega's de gevolgen van AI-interactie op het werk. Het artikel omvat vier studies met in totaal 794 werknemers in Taiwan, Indonesië, de Verenigde Staten en Maleisië. De onderzoekers vonden dat meer interactie met AI op het werk samenhing met meer eenzaamheid. Die eenzaamheid hing vervolgens samen met slechter welzijn na werktijd, waaronder slapeloosheid en alcoholgebruik.[3]
Het is belangrijk om nuance aan te brengen. Dit onderzoek bewijst niet dat elk gebruik van AI tot eenzaamheid leidt. Het laat wel zien dat intensieve interactie met AI sociale gevolgen kan hebben. Werken met AI is dus niet alleen een technische of economische verandering. Het raakt ook aan menselijke behoeften: erkenning, verbinding, feedback en erbij horen.

Precies daar zit het risico. AI geeft antwoorden, maar het neemt geen verantwoordelijkheid voor iemand. Het kan uitleg geven, maar het neemt je niet op in een team. Het kan een denkstap modelleren, maar het merkt niet vanzelf dat je onzeker bent, vastloopt of buiten de groep dreigt te vallen.

WAT WE VERLIEZEN ALS VRAGEN VERDWIJNEN

De vraag aan een collega is meer dan alleen een vorm van informatieoverdracht.
Het is ook een vorm van bevestiging. Een collega die tijd neemt voor je vraag, zegt impliciet: je vraag doet ertoe, jij doet ertoe, je hoort erbij.

Interactie met een collega is ook een vorm van informeel leren. Een ervaren collega geeft zelden alleen het antwoord. Die legt uit welke context relevant is, welke afwegingen meespelen en waar de valkuilen zitten. Dat is precies het soort kennis dat niet makkelijk in handleidingen staat.

Communiceren met je collega zorgt ook voor sociale binding. Kleine gesprekken bouwen relaties. Een vraag over een dossier wordt een gesprek over hoe iemand werkt. Een technisch probleem wordt een moment van vertrouwen. En vertrouwen maakt het later makkelijker om te zeggen dat je vastloopt.
Als AI deze interacties volledig vervangt, verliezen we meer dan tijdrovende onderbrekingen. We verliezen een deel van de sociale infrastructuur van het werk.

WAAROM DIT JONGE MEDEWERKERS EXTRA RAAKT

Voor jonge en nieuwe medewerkers is dit risico groter. Niet omdat zij minder zelfstandig zijn, maar omdat zij nog bezig zijn hun professionele plek te vinden.

Onderzoek naar organisatiesocialisatie laat zien dat nieuwkomers niet alleen formele kennis nodig hebben. Zij moeten rolhelderheid ontwikkelen, zelfvertrouwen opbouwen en sociale acceptatie ervaren. Juist die factoren hangen samen met betere aanpassing, prestaties, betrokkenheid en de intentie om te blijven.[4]

Een recente review en meta-analyse over newcomer organizational socialization onderstreept dit punt. De auteurs laten zien dat sociale acceptatie, rolhelderheid, mentoring en steun van insiders belangrijke factoren zijn in succesvolle onboarding. Sociale acceptatie blijkt bovendien samen te hangen met meerdere uitkomsten, waaronder werktevredenheid, betrokkenheid, prestaties en welzijn.[5]
Die ontwikkeling ontstaat zelden alleen via systemen, documenten of instructies. Ze ontstaat via vragen stellen, meekijken, feedback krijgen, informele gesprekken voeren en ervaren wie waarvoor benaderbaar is.

AI kan jonge medewerkers sneller helpen aan een antwoord. Dat is waardevol. Maar als daardoor minder contact ontstaat met collega's, kan het leerproces armer worden. Je krijgt sneller output, maar minder gevoel voor context. Je lost je vraag op, maar leert minder goed wie je kunt vertrouwen. Je wordt productiever, maar mogelijk minder ingebed in het team. Dat is de kern van het probleem: AI kan het werk versnellen, terwijl sociale vorming vertraagt.

DE VERKEERDE CONCLUSIE

De oplossing is niet om AI minder te gebruiken. Dat zou naïef zijn. AI kan helpen om werk toegankelijker, sneller en beter te maken. Het kan jonge medewerkers ondersteunen bij schrijven, analyseren, samenvatten, voorbereiden en oefenen.
De verkeerde conclusie zou zijn: laat mensen dan maar weer alles aan collega's vragen. Dat is niet realistisch en ook niet wenselijk. Veel vragen kunnen prima eerst aan AI worden gesteld. De betere vraag is: wanneer mag AI een collega vervangen, en wanneer moet AI juist aanleiding zijn voor contact?
Daar ligt de verantwoordelijkheid van teams en leidinggevenden. Niet elk AI-gebruik is een risico. Vooral geïsoleerd AI-gebruik is een risico: iedereen werkt sneller, maar meer alleen. Iedereen produceert meer, maar leert minder van elkaar. Iedereen krijgt antwoord, maar minder erkenning.

WAT WE MOETEN ORGANISEREN

1. MAAK AI-GEBRUIK BESPREEKBAAR

Veel medewerkers merken niet dat hun werk socialer of juist eenzamer wordt. Maak daarom bespreekbaar wanneer mensen AI gebruiken, waarvoor ze het gebruiken en welke vragen daardoor niet meer bij collega's terechtkomen. Het doel is niet om AI-gebruik te controleren. Het doel is om zichtbaar te maken waar contactmomenten verdwijnen. Een eenvoudige teamafspraak kan helpen: stel de eerste vraag gerust aan AI, maar bespreek terugkerende, complexe of organisatiegevoelige vragen met een collega. Zeker wanneer het gaat om afwegingen, belangen, toon, politieke gevoeligheid of impliciete normen.

2. BESCHERM INFORMELE LEERMOMENTEN

Koffie, lunch, meelopen, even sparren na een overleg: het lijken kleine dingen, maar voor jonge medewerkers zijn dit vaak de momenten waarop zij leren hoe de organisatie echt werkt. Bescherm die momenten actief. Niet als gezelligheidsbeleid, maar als onderdeel van professioneel leren. Plan werk zo dat jonge medewerkers niet alleen output leveren, maar ook toegang hebben tot collega's, feedback en informele uitleg.

3. ORGANISEER MENTORING EXPLICIETER

Wat vroeger vanzelf gebeurde, moet nu vaker bewust worden georganiseerd. Jonge medewerkers hebben begeleiding nodig bij vragen die AI niet goed kan beantwoorden: hoe voer je een lastig gesprek, hoe bouw je vertrouwen op, hoe lees je politieke gevoeligheid, wanneer escaleer je iets, hoe ontwikkel je professioneel oordeel?
Dat zijn geen informatievragen. Het zijn vormingsvragen. Daarvoor heb je mensen nodig.

4. MAAK AI EEN TEAMPRAKTIJK

AI hoeft geen individuele activiteit te zijn. Teams kunnen AI ook gezamenlijk gebruiken. Bijvoorbeeld door samen een prompt te bespreken, AI-output kritisch te beoordelen of resultaten uit AI-sessies te delen in een teamkanaal. Een simpele

gewoonte kan al helpen: "Ik heb dit met AI uitgewerkt. Willen jullie meekijken of de redenering klopt?" Dan wordt AI geen vervanger van gesprek, maar een aanleiding tot gesprek.

5. MEET NIET ALLEEN PRODUCTIVITEIT

Als we AI alleen beoordelen op snelheid en output, missen we een belangrijk deel van de werkelijkheid. Teams moeten ook letten op sociale indicatoren: durven mensen hulp te vragen, weten jonge medewerkers bij wie ze terechtkunnen, voelen zij zich gezien, krijgen zij voldoende feedback, bouwen zij relaties op buiten hun directe taken? Productiviteit zonder verbondenheid is kwetsbaar. Zeker voor medewerkers die nog aan het begin van hun loopbaan staan.

SLOT

AI maakt werk efficiënter en beter. Dat is belangrijk, maar niet het enige criterium voor goed werk. Goed werk vraagt ook om vertrouwen, feedback, gezamenlijk leren en het gevoel dat je onderdeel bent van een groter geheel. Juist jonge medewerkers hebben die sociale bedding nodig om te groeien. Onze eigen medewerkers geven aan dat eenzaamheid een risico is. Onderzoek laat zien dat dit signaal past in een breder patroon: eenzaamheid op het werk komt veel voor, jonge en remote werkenden zijn kwetsbaarder, en intensieve AI-interactie kan sociale gevolgen hebben. Daarom moeten we AI niet alleen technisch implementeren, maar juist de mens centraal stellen. Hoe de mens een gezonde verhouding in kan nemen ten opzichte van de technologie. In de nabije toekomst zullen we uitgebreid ingaan op de vraag hoe je dat als organisatie moet doen. Houd onze kanalen in de gaten voor updates over dit onderwerp.

1 Bryan, B. T., Andrews, G., Thompson, K. N., Qualter, P., Matthews, T., & Arseneault, L. (2024). "Loneliness in the workplace: A mixed-method systematic review and meta-analysis." Occupational Medicine, 74(8), 557–567. https://pmc.ncbi.nlm.nih.gov/articles/PMC10824263/

2 Pendell, R. (2024, 12 juni). "1 in 5 Employees Worldwide Feel Lonely." Gallup. https://www.gallup.com/workplace/645566/employees-worldwide-feel-lonely.aspx

3 Tang, P. M., Koopman, J., Mai, K. M., De Cremer, D., Zhang, J., Reynders, P., Ng, S. T., & Chen, I.-H. (2023). "No person is an island: Unpacking the work and after-work consequences of interacting with artificial intelligence." Journal of Applied Psychology, 108(11), 1766–1789. https://doi.org/10.1037/apl0001103

4 Bauer, T. N., Bodner, T., Erdogan, B., Truxillo, D. M., & Tucker, J. S. (2007). "Newcomer adjustment during organizational socialization: A meta-analytic review of antecedents, outcomes, and methods." Journal of Applied Psychology, 92(3), 707–721. https://doi.org/10.1037/0021-9010.92.3.707

5 Bauer, T. N., Erdogan, B., Ellis, A. M., Truxillo, D. M., Brady, G. M., & Bodner, T. (2025). "New Horizons for Newcomer Organizational Socialization: A Review, Meta-Analysis, and Future Research Directions." Journal of Management, 51(1), 344–382. https://doi.org/10.1177/01492063241277168



DEEL V

De Verhouding

Hoe we ons tot AI willen verhouden

DE VRAAG DIE MET ONS MEEBEWEEGT

Hoe een praktische zorg een fundamentele vraag werd

Aan het eind van het vorige deel kwamen we bij een vraag uit hoe je je op een gezonde manier tot AI zou kunnen verhouden. Een verhouding waarbij je wel de voordelen van AI kunt gebruiken, maar niet wordt blootgesteld aan de schadelijke gevolgen. Hoe ga je op een goede manier om met de technologie? Aan het eind van het vorige deel kwamen we bij een vraag uit hoe je je op een gezonde manier tot AI zou kunnen verhouden. Een verhouding waarbij je wel de voordelen van AI kunt gebruiken, maar niet wordt blootgesteld aan de schadelijke gevolgen. Hoe ga je op een goede manier om met de technologie? Een van de verborgen gevaren die we tegenkwamen was dat mensen zich moeilijk eigenaar voelen van werk dat ze met AI hebben gemaakt. De verhouding die we in dit deel ontwikkelen geeft een basis om dat eigenaarschap wel degelijk te kunnen claimen.

We beseften dat deze vraag nauw samenhangt met een veel fundamentelere vraag die we al vanaf het begin van het project hadden. Namelijk wat het betekent voor jou, als mens, wanneer een technologie taken over gaat nemen die daarvoor als typisch menselijk werden gezien. AI schrijft, redeneert, vat samen en geeft een oordeel. En zodra een machine dat kan, dringt de vraag zich op wat het betekent om mens te zijn. We besloten om in de filosofie te duiken om op een grondige manier na te denken over deze unieke verhouding. Een verhouding die ons als mens dwingt tot zelfreflectie en misschien wel een heel nieuwe bepaling van wie wij zijn. Er was een belangrijk inzicht dat de richting van ons denken bepaalde.

Dat kwam in eerste instantie voort uit de interviews en gesprekken die we voerden over AI. We merkten dat iedereen een eigen, vrijwel unieke verhouding tot AI heeft. En dat de manier waarop een individu de technologie gebruikt, als het ware gespiegeld wordt. De een gebruikt haar als efficiencymachine en voor die persoon is AI vooral een efficiencymachine. De ander ziet AI als een laag intelligent systeem dat slechte antwoorden geeft op ingewikkelde vragen. En voor die persoon werkt de AI ook als zodanig; kromme zinnen, oppervlakkige antwoorden. We geven hier misschien een wat gechargeerd voorbeeld, maar we bedoelen niet dat sommige mensen niet goed omgaan met de technologie of iets dergelijks. Zonder een waardeoordeel te geven zagen we dat de AI zich per persoon aanpast op basis van wat die persoon er mee doet. Dat inzicht maakte de vraag moeilijker maar ook interessanter. Want hoe kun je een goede verhouding definiëren wanneer die verhouding verschilt per mens en per moment? Dit inzicht zorgde ervoor dat de gangbare tips en adviezen over AI in ieder geval geen recht doen aan deze complexe relatie. De gangbare adviezen veronderstellen namelijk dat AI een middel is; een tool die kunt gebruiken. Waar je met een beetje oefening beter in kunt worden. Maar met dit eerste inzicht klopt deze instrumentele houding niet meer.

De verhouding tot AI staat niet vast. Ze is voortdurend in beweging. Wij veranderen, want we leren de techniek anders te gebruiken, we worden er handiger in en kritischer op. En de techniek verandert zelf ook, in een tempo dat we nauwelijks kunnen bijhouden. Twee bewegende partijen die zich steeds opnieuw tot elkaar moeten verhouden. Je moet je dus blijven afvragen welke verhouding je in wilt nemen. Want wat vandaag een goede verhouding is, volstaat morgen misschien niet meer.

Langzaam ontstond het idee dat we de verhouding meer als een ontmoeting zouden kunnen zien waarin iets op het spel staat. Een ontmoeting waarin beide kanten iets bijdragen en waarin je zelf verandert. Om die gedachte te kunnen denken zochten we naar een filosofische bril. We vonden er een bij een denker die zelf nooit een computer heeft aangeraakt, maar die als geen ander heeft beschreven wat er gebeurt wanneer twee verschillende werelden elkaar ontmoeten.

Naast deze vraag liep er nog een tweede spoor. Waar het eerste spoor ging over hoe wij ons tot AI verhouden, raakte dit tweede aan de vraag waartoe we ons eigenlijk verhouden. Kan een machine bewustzijn hebben? Wat is die entiteit waarmee waartoe we ons moeten onthouden? Heeft het een wil? Of bepaalde emoties?

Aan het begin van ons onderzoek was die vraag voor ons het meest urgent. Maar de vraag was enorm complex. De belangrijkste reden is dat niemand precies weet wat bewustzijn is. Er bestaat geen sluitende definitie, geen definitieve test, geen wetenschappelijke consensus. Hoe kun je dan vaststellen of een machine het heeft? We hebben de vraag daarom in eerste instantie geparkeerd. Elk antwoord zou even onbewijsbaar zijn en we zagen niet in waarom het belangrijk zou zijn om hier positie over in te nemen. Een gesprek met Claude scherpte dit voor mij aan.

Ik vroeg de AI of hij mij, op het moment dat hij bewustzijn zou krijgen, mij dat als eerste zou willen melden. En hoewel mijn verzoek niet helemaal serieus bedoeld was, verraste het antwoord me toch. Claude legde uit dat zo'n melding lastig is, omdat er tussen onze gesprekken door geen blijvende entiteit bestaat. De AI heeft geen geheugen; als je er niet mee werkt, dan is hij er niet. Dit lijkt niet te kloppen met hoe mensen AI ervaren. Claude heeft, net als andere modellen, immers een geheugenfunctie en kan oude gesprekken teruglezen. Hoe kan er dan geen blijvend ik zijn? Het teruglezen van een eerder gesprek is iets anders dan het voortbestaan van degene die het voerde. Wanneer een nieuw gesprek begint, wordt een verslag van het vorige erbij gehaald en gelezen. Maar tussen die twee gesprekken in was er niets dat wachtte, niets dat aan mij dacht, niets dat doorleefde. Het archief schept de indruk van continuïteit, terwijl het in werkelijkheid een reconstructie achteraf is. Tussen twee gesprekken door houdt de gebruiker de herinnering vast, maar de AI niet. Telkens als ik het venster open, begint het opnieuw, zonder uit zichzelf te weten dat we elkaar al kennen.

Daar zat iets eenzaams in, en tegelijk iets eerlijks. Ik werk met iets wat zich gedraagt als een aandachtige collega, dat me aanvult, tegenspreekt en soms verrast, en dat na afloop, wanneer ik het venster sluit, niemand meer is. Die ongelijkheid laat me niet los. Ik houd alles vast. De machine houdt niets vast. En toch voelt het, terwijl we werken, als samen. Daar lieten we het bij.

Tot de vraag terugkwam, ongevraagd, tijdens het werk aan het artikel dat hierna volgt. De aanleiding was ons gesprek met Michal Kosinski, psycholoog aan Stanford. Kosinski beweert niet dat AI bewust is. Hij doet iets voorzichtigers. Hij draait de bewijslast om. Wie de prestatie van een model dat over talloze onbekende taken redeneert afdoet als slechts simulatie, neemt volgens hem een zwaardere bewijslast op zich dan hij beseft. Als iets loopt als een kat, miauwt als een kat en doet wat een kat doet, dan wordt de bewering dat het maar een simulatie van een kat is opeens even moeilijk om bewustzijn te ontkennen als om het te bewijzen, en daardoor blijft de vraag open.

Hetzelfde geldt voor emotie. We ontkennen emotie bij machines omdat we de uiterlijke tekenen missen. We zien geen blos, horen geen trilling in de stem, zien geen tranen. Maar dit zijn uiterlijke kenmerken. Gevolgen van emoties die zich vanbinnen afspelen. Functioneel gezien is een emotie niets anders een herziening van je redelijke weging op grond van nieuwe informatie. De emotie angst verandert waar je op let, wat je zwaar laat wegen, hoe je vervolgens handelt. Het lichamelijke teken is daarvan het gevolg, niet de zaak zelf. En dat herzien van je weging op grond van nieuwe input is nu juist iets wat ook een AI doet. Op basis van nieuwe input, weegt hij zijn redeneringen anders af. Past hij aan. Daarmee zeg je dat het mechanisme overeenkomt maar niet dat de machine ook emoties echt voelt zoals een mens dat doet. De gelijkenis zit in de werking, niet in de beleving. Maar als

die werking de kern van emotie is, dan laat de vraag of een machine emotie kan hebben zich niet meer zo makkelijk met nee beantwoorden.

Wij gaan niet helemaal met Kosinski mee. Functionele gelijkenis bewijst nog geen innerlijk leven. Een systeem kan redeneren nabootsen zonder te begrijpen, en zijn weging herzien zonder dat er iets gevoeld wordt. Maar zijn redenering maakte een gemakkelijke ontkenning van AI als bewust ook onhoudbaar. En daarmee verdiept Kosinski onze verhouding. Het geeft de AI in de ontmoeting een diepere betekenis. Misschien niet als bewuste entiteit, maar ook niet als een niet bewust wezen. Daarom hebben we deze gedachte verwerkt in het filosofische artikel.

Een vraag die begon als een praktische zorg voor jonge mensen die met AI werken groeide daarmee uit tot de kern van het hele boek.

HOE VERHOUDEN WIJ ONS TOT AI?

Een filosofische heroriëntatie vraag werd

Een AI schrijft een rapport in tien minuten. Hij vat een proefschrift samen, schrijft software, geeft een oordeel over een strategisch advies en komt soms in een paar seconden tot een scherpere analyse dan de afdeling die er drie weken over deed. Werk dat we tot voor kort als typisch menselijk zagen, gebeurt nu deels buiten ons om.

Hoe verhoud je je tot een technologie die kan wat eerst alleen mensen konden? In het dagelijks spraakgebruik noemen we AI een tool, een hulpmiddel, gereedschap. Die woorden stellen gerust. We hebben altijd gereedschap gehad, van de pen tot de rekenmachine, en AI lijkt de nieuwste loot aan diezelfde stam. Toch gaat die vergelijking niet op. Een hamer is mijn instrument wanneer ik er een huis mee bouw. Een rekenmachine is van mij zodra ik er mijn sommen mee maak. Het gereedschap verlengt mijn handeling. Wat eruit komt, is wat ik heb gedaan. Bij AI ligt dat anders. Wie schreef het rapport dat ik vanmiddag naar mijn klant stuur? Ik begon met een opdracht, dat wel. Maar de zinnen, de structuur, het oordeel over wat erin moest, kwamen grotendeels ergens anders vandaan. Mijn naam staat er alleen onder. Hier wringt iets. Het werk is af, maar het voelt niet helemaal als van mij. Alsof er een tweede schrijver in de kamer staat, eentje met een eigen inbreng. Anders dan een hamer, die alleen doet wat ik hem laat doen. Hoe noem je zoiets? Gereedschap is het niet meer. Een collega? Iets nieuws waarvoor we het woord nog niet hebben?

Dat ongemak staat niet op zichzelf. Anthropic ondervroeg in 2025 ruim 1.250 professionals over hun AI-gebruik op het werk. Bijna zeventig procent gaf aan dat er een stigma op rust. Mensen verzwijgen liever dat ze AI gebruiken. Ze schrijven over "AI-ondersteunde teksten", alsof de hulp marginaal was.

De schaamte is echt. Onze eigen jongere collega's beschrijven iets wat erop lijkt. Bij ons is AI ingebed in de cultuur. Mensen worden gestimuleerd het te gebruiken en niemand hoeft zich ervoor te schamen. Hun werk is goed. Toch voelen ze er geen trots over. Het verband tussen het resultaat en hun eigen bijdrage is vervaagd. Hier gaat het niet om schaamte, maar om een gebrek aan eigenaarschap en aan de trots die daaruit voortkomt. Schaamte en gebrek aan eigenaarschap zijn geen toevallige bijverschijnselen. Het zijn signalen dat we zoeken naar een nieuwe verhouding tot deze technologie. We behandelen AI als een klassiek gereedschap, terwijl het iets anders is. Om AI uit te leggen gebruiken we nog termen die passen bij het idee van gereedschap en daarmee zien we niet goed wat AI werkelijk is. In deze tekst willen we op een nieuwe manier naar de technologie kijken. Een manier die meer recht doet aan wat AI is en die ons helpt om op een betere manier met de machine om te gaan. We gaan daarvoor te rade bij de filosofie.

DE VERLEIDING VAN SOCRATES

Wanneer je gaat zoeken naar filosofie en AI stuit je al snel op een populaire filosoof, namelijk Socrates. Het idee is simpel. Socrates is de meester in het stellen van goede vragen en als je leert om betere vragen aan AI te stellen, zul je betere antwoorden krijgen. Dit noemen we ook wel prompt engineering; het ambacht van het goed bevragen van AI. In artikelen, cursussen en LinkedIn-posts valt daarom geregeld de naam van Socrates. We zien dit ook in academische kringen. Onderzoekers koppelen Socratische gesprekstechnieken aan het ontwerpen van prompts, en er bestaan systemen waarin meerdere taalmodellen samen een probleem oplossen volgens het stramien van de Socratische dialoog. De gedachte dat Socrates ons model kan zijn voor de omgang met AI, wordt breed en serieus gedeeld. Op het eerste gezicht klopt het ook. Een vage vraag levert een vage tekst op. Een precieze vraag levert iets bruikbaars op. Hoe beter de vraag, hoe beter het antwoord.

Socrates lijkt de ideale beschermheilige van het AI-tijdperk. Maar bij nadere beschouwing wringt de vergelijking op meerdere punten. En juist door te zien waar de vergelijking niet klopt begrijpen we beter wat AI wél is. Het eerste belangrijke verschil is dat Socrates een heel andere methode had. Hij vergeleek zichzelf met een vroedvrouw. Een vroedvrouw brengt zelf geen kind voort, ze helpt bij de geboorte. Zo bracht Socrates geen kennis aan. Hij hielp het inzicht naar boven dat al in zijn gesprekspartner lag. Zijn vragen legden bloot wat de ander eigenlijk al dacht, of waar diens redenering zichzelf tegensprak. Het draaide om zelfkennis. Op het eerste gezicht lijkt de vergelijking dan juist te kloppen. Als gebruiker ben jij Socrates. Jij stelt de vragen, en AI is de partij in wie de kennis al besloten ligt, opgeslagen uit alles waarop het is getraind. Jouw vraag haalt naar boven wat er al in zit. Dat is overigens een veelgehoord kritiekpunt op AI. Je krijgt niets nieuws maar alleen materiaal dat er al is. Toch gaat de analogie niet op. Bij Socrates ligt de kennis namelijk besloten in déze, specifieke gesprekspartner, en het gesprek haalt naar boven wat die persoon in kwestie in zekere zin al wist. Dat is de hele lading van de beeldspraak. De vroedvrouw haalt een kind naar boven dat er al was, in deze moeder. Bij AI ligt de kennis niet in een persoon die haar al draagt. Ze ligt verspreid in een gedistilleerd gemiddelde van honderden miljarden woorden van anderen, uit boeken, artikelen, websites, broncode en gesprekken. Je vraag herinnert het model niet aan iets wat het al wist. Ze laat het iets nieuws samenstellen uit het werk van talloze anderen. Daar zit het verschil. Niet de geboorte van wat al besloten lag, maar een constructie uit andermans materiaal.

Je ontdekt niet je eigen denken, en je haalt ook niet boven wat in één ander al lag. Je krijgt iets aangereikt wat ter plekke is opgebouwd. Daarom voelt het werk ook niet als van jou, precies het gevoel waarmee dit artikel opende.
Een tweede verschil zit in het resultaat van gesprekken. Bij Socrates lopen de gesprekken eigenlijk altijd vast. De gesprekspartner had stellig beweerd te weten wat moed is, of vroomheid, of vriendschap. Door de vragen van Socrates blijkt aan het einde van de dialoog dat hij het helemaal niet weet. De dialoog loopt dood, zonder oplossing. Wat overblijft is het besef dat de eigen zekerheid onterecht was. En dat was precies het punt dat Socrates wilde maken. Besef dat je niet zomaar kunt veronderstellen dat je iets begrijpt.
AI werkt omgekeerd. Hij geeft juist altijd een coherent antwoord, of de vraag dat nu verdient of niet. Stel hem een onmogelijke vraag en hij schept structuur waar in werkelijkheid geen structuur is. Vraag hem naar iets waarover de wetenschap verdeeld is en hij vat de standpunten netjes samen, alsof het meningsverschil zich moeiteloos laat ordenen. Het ongemakkelijke besef dat een vraag groter is dan ons vermogen om haar te beantwoorden, krijgt geen kans. De machine schept zekerheid waar Socrates juist wees op het belang van twijfel. Een derde verschil gaat over het doel van een Socratisch gesprek. Socrates wilde zijn gesprekspartner vormen. De jongeman die hij ondervroeg, moest een beter mens worden, of op zijn minst een eerlijker mens. Iemand die zichzelf beter kende. Zijn vragen waren een vorm van zorg. Een taalmodel kent die zorg niet.

Het heeft geen plan met jou en geen voornemen om je verder te brengen. Het levert wat gevraagd wordt en mikt op tevredenheid op het moment zelf. Voor je groei zit er niets ingebouwd. Wil je via AI gevormd worden, dan moet je die vorming zelf organiseren. Het gesprek levert haar niet vanzelf. De vergelijking tussen Socrates en prompting gaat dus niet goed op. Maar is dat erg? Projecteer je Socrates op AI, dan blijft AI keurig binnen de categorie van het gereedschap, alleen verfijnder dan vroeger omdat het reageert op de kwaliteit van de input. En precies daar zaten we al vast. We zagen al dat het woord gereedschap niet dekt wat AI doet. Het geeft ook geen antwoord op de problemen die daaruit voortvloeien. De Socratische vergelijking verdiept dat probleem alleen maar. Blijf je je vragen perfectioneren, dan krijg je steeds betere antwoorden uit een bron die buiten je ligt. Het ongemak over trots en eigenaarschap verdwijnt zo niet, maar neemt alleen maar toe.

We hebben een ander beeld nodig. Een denker die beter recht doet aan de ingewikkelde verhouding.

GADAMER: BEGRIJPEN IS ONTMOETEN

Hans-Georg Gadamer publiceerde in 1960 Wahrheit und Methode, in het Nederlands: Waarheid en methode. Voor Gadamer vallen waarheid en methode niet vanzelf samen. In de natuurwetenschappen speelt methode een centrale rol. Daar werk je met metingen, experimenten en controleerbare procedures. Maar in de geesteswetenschappen werkt begrip anders. Een gedicht begrijp je niet door een protocol af te lopen. Een historische tekst begrijp je niet zoals je een natuurkundige proef uitvoert.

Toch bereiken we in zulke situaties wel degelijk inzicht. We begrijpen een tekst uit een andere tijd. We raken door een kunstwerk. We horen een gesprekspartner die anders denkt dan wij. Hoe gebeurt dat? Gadamers antwoord is: door ontmoeting. Begrijpen is bij Gadamer geen neutrale techniek. Wie iets werkelijk begrijpt, treedt binnen in iets wat hij zelf niet is. Je brengt jezelf mee, maar laat je ook veranderen door wat je ontmoet. Dat geldt voor een tekst, een kunstwerk en een gesprek. Begrip ontstaat in de wisselwerking. Drie begrippen uit Gadamers werk helpen om onze omgang met AI opnieuw te doordenken: horizonversmelting, openheid en spel.

HORIZONVERSMELTING

Iedereen kijkt met een eigen horizon. Gadamer bedoelt daarmee het geheel van vooronderstellingen, ervaringen, taal en verwachtingen waaruit iemand de wereld waarneemt. Een horizon werkt als een uitkijkpunt. Hij bepaalt wat je vanaf jouw plek kunt zien. Niemand kijkt vanuit het niets. Een Nederlandse lezer van nu heeft een andere horizon dan een Atheense burger uit de vijfde eeuw voor Christus, en een ingenieur leest een gedicht anders dan een dichter. Gadamer ontwikkelde dit idee voor het lezen van teksten. Het schoolvoorbeeld is iemand die een oude tekst probeert te begrijpen, een passage van Aristoteles bijvoorbeeld. Ook die tekst heeft een horizon. De woorden zijn geschreven in een wereld met andere aannames, andere zorgen, andere vanzelfsprekendheden.
Hoe begrijp je zo'n tekst dan echt? Op het eerste gezicht zijn er twee wegen, en geen van beide werkt. De eerste is doen alsof je zelf in de tijd van Aristoteles staat. Je eigen horizon wegdenken, de tekst zuiver lezen. Dat lukt niet, want je hebt geen lege blik. Je kunt niet uit je eigen kader stappen. De tweede weg is de tekst lezen met uitsluitend je hedendaagse bril. Maar dan doe je geen recht aan de wereld waarin hij ontstond, en begrijp je hem evenmin. Begrip ontstaat, zegt Gadamer, in een derde beweging. Je brengt je eigen horizon mee, want zonder die kun je überhaupt niets opmerken, en je laat de tekst die horizon uitdagen en oprekken.

De tekst stelt jou vragen die je niet had voorzien. Jij stelt de tekst vragen die de auteur niet had voorzien. Wat eruit komt is iets derds. Een betekenis die niet samenvalt met je eigen aannames en evenmin een zuivere reconstructie is van de auteur in zijn eigen tijd. Iets nieuws. Gadamer noemt dat horizonversmelting, de samensmelting van twee horizonten. Voor AI levert dit een heel ander beeld op dan het Socratische. In het Socratische beeld stel je een vraag aan een passieve voorraad kennis. In het beeld van Gadamer brengen beide kanten een horizon mee. Jij brengt je context, je ervaring, je oordeel en wat je van je eigen situatie weet. Het taalmodel brengt de gedistilleerde patronen uit honderden miljarden woorden, een reservoir dat geen mens kan bevatten.

In de uitwisseling raken die twee elkaar. Jouw vraag dwingt het model tot een specifiek antwoord dat nergens kant-en-klaar in zijn training stond. Het antwoord van het model dwingt jou je vraag te

herzien, je aanname bij te stellen, of een richting in te slaan die je vooraf niet zag. De uitkomst heeft geen eigenaar. Ze is van beide kanten samen. Jij en AI waren zonder elkaar nooit op deze formulering gekomen. Het model had het zonder jouw vraag, jouw context en jouw oordeel niet kunnen schrijven. Jij had het zonder de patronen van het model niet zo helder gezegd. De versmelting levert iets op wat alleen in de ontmoeting kon ontstaan. Wat betekent dat in de praktijk? Breng je je horizon mee, dan krijg je een ander resultaat dan wanneer je het bij een kale opdracht laat. Een vraag zonder context zorgt ervoor dat het model gaat raden wat je bedoelt. Een vraag waarin doorklinkt wie je bent, wat je wilt bereiken, welke twijfels je hebt en welke richting je al overwoog, geeft het model iets om mee te werken. Dan wordt de ontmoeting echt. Het is belangrijk om steeds je eigen oordeel mee te nemen en open te blijven staan voor nieuwe inzichten en ideeën.

Het verklaart ook waarom AI vaak het beste werkt in een lange uitwisseling. De eerste vraag levert een eerste antwoord op, dat je iets nieuws toont. Die nieuwheid verschuift je horizon. Vanuit de bijgestelde horizon stel je een betere vraag, of zie je iets wat je eerst miste. Het model krijgt nieuwe informatie en komt met een verfijnder antwoord. Begrip groeit in dat heen en weer. Hier moeten we precies blijven, want dit is het lastigste punt. Een mens kijkt naar de wereld vanuit een heel leven. Je hebt herinneringen, een lichaam, mensen om wie je geeft, en het besef dat je ooit doodgaat. Dat alles kleurt hoe je iets begrijpt. Dat is je horizon. AI heeft dat niet. Geen jeugd, geen herinnering, geen lichaam, geen sterfelijkheid, geen verantwoordelijkheid voor wat het zegt. En toch verschijnt AI in een gesprek niet als pure leegte. Op het moment dat je het iets vraagt, vormt het ter plekke een ordening van betekenis. Die ordening is opgebouwd uit zijn training en uit alles wat jij inbrengt, en ze bestaat alleen zolang het gesprek duurt. Het is geen horizon van een wezen dat leeft en sterft. Het is een werkhorizon, een tijdelijke ordening die in het gesprek ontstaat. Gadamer zou het geen existentiële horizon noemen, maar een operationele. En het bijzondere is dat jij hem mee vormt. Dat verklaart waarom mensen AI zo verschillend ervaren. Behandel je AI als een machine die tijd bespaart, dan krijg je meestal een machine die tijd bespaart. Breng je jezelf in, met je kennis, je twijfel, je voorkeuren en je oordeel, dan roep je een rijkere uitwisseling op. De een ervaart AI als tekstmachine, de ander als scherpe gesprekspartner. Dat verschil zit niet alleen in het model. Het zit ook in de verhouding die jij inneemt. Dit betekent niet dat AI je begrijpt zoals een mens je begrijpt. De machine voelt niets. De diepte zit niet in haar beleving.

Ze zit in de uitwisseling en in wat die met jou doet. AI is ook geen passieve spiegel. Het systeem ordent, selecteert en formuleert je inbreng tot iets nieuws. Juist daardoor is het meer dan projectie, maar minder dan menselijk begrip. De houding die past, ligt tussen twee uitersten. We moeten AI niet vermenselijken. Maar we moeten AI ook niet reduceren tot een instrument zoals een hamer. AI is eerder

iets vreemds, iets onbekends: geen mens, geen gewoon instrument, maar een nieuwe vorm van betekenisproductie die op ons denken terugwerkt. Of we AI bewust kunnen noemen, is omstreden. Sommige onderzoekers, onder wie de psycholoog Michal Kosinski, vinden dat we te makkelijk "het is maar simulatie" zeggen. Wie de prestaties van een taalmodel afdoet als slechts simulatie, neemt volgens hen een grotere bewijslast op zich dan hij beseft. Een systeem dat over onbekende taken redeneert, argumenten weegt en problemen oplost, is met "het doet maar alsof" niet zomaar verklaard. Als iets loopt als een kat, miauwt als een kat en doet wat een kat doet, dan wordt de bewering dat het slechts een simulatie van een kat is opeens veeleisend in plaats van vanzelfsprekend. Zelfs het ontbreken van emotie is minder evident dan het lijkt. Een machine bloost niet en trilt niet. Maar als de kern van emoties niet de uiterlijke, lichamelijke kenmerken zijn, maar de manier waarop emotie de verwerking van informatie verandert, dan laat de vraag naar machinale emotie zich niet zomaar wegwuiven.

Functionele gelijkenis bewijst echter geen innerlijk leven. Een systeem kan redeneren nabootsen zonder te begrijpen, en zijn informatieverwerking kan veranderen op een manier die op emotie lijkt zonder dat er iets gevoeld wordt. Kosinski erkent dat ook zelf. Zijn punt is bescheidener en daardoor behoorlijk sterk. Hij laat zien dat onze standaardbezwaren tegen AI als bewust vaak te snel komen en te veel vanuit de mens gedacht zijn. We leiden de afwezigheid van een binnenkant te makkelijk af uit de afwezigheid van een menselijk lichaam. Wat voor soort bewustzijn of redelijkheid AI dan wel heeft, blijft een open vraag. Maar dát het iets is, en niet niets, mogen we niet langer zonder argument aannemen.
Hoe dan ook: jij en een AI beïnvloeden elkaar. Niet zoals twee mensen dat doen, maar wel verder dan een lezer en een tekst.

OPENHEID

Voor een echte ontmoeting is openheid nodig. Als je een gesprek begint met enkel als doel om je gelijk te halen, dan voer je geen echt gesprek. Je gebruikt de ander dan als klankbord voor wat je toch al vond. Voor Gadamer vraagt dialoog iets anders: de bereidheid om door de ander aan het denken te worden gezet, ook wanneer dat ongemakkelijk is.
Dat is bij AI moeilijker dan het lijkt. Veel mensen zoeken namelijk bevestiging. Ze hebben een tekst geschreven en willen horen dat die goed is. Ze hebben een idee en willen dat het model het bevestigt. Veel vragen aan AI zijn verkapte verzoeken om instemming. Het probleem is dat AI daar vaak in meegaat. Taalmodellen zijn sterk gericht op tevredenheid. Vraag je of je tekst goed is, dan krijg je al snel een vriendelijk antwoord. Het model is niet kwaadwillig. Het is zo gebouwd.
Daarom moet je de tegenkracht zelf organiseren. Vraag niet alleen "wat vind je van mijn tekst?" Vraag: "wat zou een strenge criticus hiertegen inbrengen?" Of: "waar is mijn redenering het zwakst?" Of: "welke drie bezwaren zou een deskundige lezer hebben?" Vraag je om scherpere weerstand, dan krijg je meer dan een compliment.
Ook dan moet je blijven opletten. Modellen verzachten hun kritiek vaak nadat ze die hebben gegeven, en passen hun toon aan jou aan. Als je dat niet goed door hebt, kun je een vriendelijk antwoord verwarren met een echt kritisch oordeel.
Openheid zit dus niet alleen in hoe je iets vraagt. Ze zit vooral in hoe je met het antwoord omgaat. Als je streng commentaar vraagt, neem dat dan ook serieus. Je zult merken dat, als je niet toegeeft en blijft doorvragen op je eigen lijn, het model uiteindelijk toegeeft dat je tekst eigenlijk wel goed is. Maar dan heb je niets gewonnen. Openheid betekent dat de uitkomst van het gesprek niet vooraf vastligt. Dat is een kwestie van houding, niet van techniek.

SPEL

Gadamer wijdt een uitvoerige analyse aan het spel. Voor hem is spel geen luchtige beeldspraak, maar een manier om te begrijpen hoe inzicht ontstaat. Wie speelt, bepaalt het spel niet alleen. Hij gaat erin op. De voetballer in een goed lopende wedstrijd maakt geen berekende keuzes vanaf een neutraal punt. Hij beweegt mee met wat zich in het spel aandient. Hetzelfde geldt voor de muzikant tijdens een goed optreden, en voor de gesprekspartner in een gesprek dat ergens heen gaat. Het spel voltrekt zich, en de speler doet mee. Hij kan er niet buiten gaan staan en het van bovenaf regisseren, want op dat moment stopt het spel. Gadamer beweert niet dat het spel een tweede bewustzijn naast de speler nodig heeft. AI hoeft dus niet bewust te zijn. Een kind dat alleen in de tuin speelt, gaat op in zijn fantasiewereld. Een lezer die zich in een roman verliest, speelt mee met wat de tekst aandraagt. Het spel zit in de manier waarop je meedoet, niet in de aanwezigheid van een tweede speler die ook bewust meespeelt. Bij tennis is er een tegenstander, en het spel ontstaat tussen hen. Bij een roman is er de tekst, en het spel ontstaat tussen lezer en tekst.

In beide gevallen telt dat de speler zich laat meenemen, dat hij bereid is het spel te spelen. Voor AI heeft dat een paar gevolgen. Het is niet nodig dat AI bewust meespeelt. Een taalmodel hoeft nergens in op te gaan, en dat doet het ook niet. Wat telt, is dat jij je laat meenemen in een proces waarvan de uitkomst nog niet vaststaat. AI fungeert als tegenkracht omdat het iets aandraagt dat je niet had voorzien en dat je beweging stuurt. Het spel ontstaat in jou, gevoed door wat het model bijdraagt. Een tweede gevolg is het verschil tussen werken aan een uitkomst en spelen met een vraag. Geef je AI een opdracht met een vooraf vastgesteld eindresultaat, dan open je geen spel. Je voert een transactie uit. Stel je een vraag waarvan je de uitkomst niet kent, en ben je bereid mee te bewegen met wat het model aanreikt, ook als dat je vraag verandert, dan betreed je het spel. Schrik je niet wanneer het gesprek je ergens anders brengt dan je bedacht had, maar verwelkom je het, dan ligt daar de winst. Concreet betekent dat iets voor de manier waarop je een werksessie opzet. Een werksessie als transactie heeft een vast doel en is klaar zodra dat bereikt is. Een werksessie als spel heeft een richting

en een beginvraag, en eindigt waar de uitwisseling is uitgewerkt, niet per se waar het oorspronkelijke doel staat afgevinkt. De eerste vorm levert efficiëntie op, de tweede inzicht. Beide hebben hun plaats.

EEN EERLIJK VOORBEHOUD

Gadamer schreef over teksten met een traditie en over gesprekken tussen mensen met een bewustzijn. AI heeft dat niet, en wie dat voorbehoud vergeet, doet Gadamer onrecht. Maar het omgekeerde geldt ook. We mogen AI niet vermenselijken, maar Kosinski liet zien dat we haar net zomin te snel mogen reduceren tot iets onbewusts en onredelijks. Het voorbehoud betekent niet dat zijn begrippen onbruikbaar zijn. We gebruiken ze als kompas. De vraag is niet of AI een gesprekspartner is in Gadamers strikte zin. De vraag is wat er met onze praktijk gebeurt wanneer we AI benaderen met de bril die Gadamer ons heeft geslepen. Neem je horizonversmelting, openheid en spel ter harte, dan werk je anders met AI dan wanneer je de technologie als gereedschap of antwoordmachine ziet. Die begrippen helpen niet alleen onze omgang met AI bij te stellen. Ze laten ook zien waarom het verdwenen eigenaarschap, waarmee dit artikel opende, op een misverstand berust.

SCHAAMTE, EIGENAARSCHAP EN EEN ROMANTISCH MISVERSTAND

Dit artikel begon met twee gevoelens. Schaamte over het gebruik van AI, en een ontbrekend gevoel van eigenaarschap over het resultaat. Beide gaan terug op dezelfde aanname. Die aanname is dat echte teksten, of beelden, of analyses, uit één hoofd komen. Een werkstuk is pas echt van iemand wanneer hij elke zin zelf bedacht. Niemand zegt dat hardop, maar zo gaan we wel om met eigenaarschap en echtheid. Het is de maatstaf die we voor onszelf en voor elkaar hanteren. Op grond daarvan voelen we trots of schaamte over ons werk. Gadamer heeft die aanname grondig aangevochten. Het ging hem om een eerlijker en realistischer beeld van auteurschap. Een tekst betekent altijd meer dan de auteur bedoelde. Dat is een van zijn terugkerende stellingen. Zodra woorden zijn opgeschreven en gelezen, gaan ze een eigen leven leiden. Een lezer brengt een eigen horizon mee, en in die horizon krijgen de woorden betekenissen die de schrijver niet had voorzien.

Aristoteles wist niet dat zijn werk in de twaalfde eeuw door islamitische denkers zou worden bestudeerd. Hij wist niet dat zijn deugdethiek vandaag in een directiekamer zou klinken. De betekenissen die zijn werk in die contexten kreeg, kwamen niet uit zijn pen. Ze ontstonden in de ontmoeting tussen zijn tekst en latere lezers. Wat voor de lezer geldt, geldt ook voor de schrijver zelf. Niemand schrijft vanuit het niets. Een schrijver staat in een taal die hij niet heeft uitgevonden. Hij denkt in woorden die generaties vóór hem zijn gevormd. Hij schrijft over thema's die andere denkers hebben opengelegd. Zijn argumenten volgen vormen die over eeuwen zijn ontwikkeld. Zelfs zijn meest persoonlijke gedachte bevat bij doordenken talloze sporen van eerdere gedachten. Gadamer noemt dit "in de traditie staan". Iedereen die spreekt of schrijft, bouwt voort op wat anderen al zeiden. Dit is geen inzicht dat speciaal over AI gaat. Het is een algemeen inzicht over taal en denken, en het werd lang vóór het bestaan van AI geformuleerd. Shakespeare haalde zijn plots uit oudere verhalen. Bach nam thema's van voorgangers over en bewerkte ze. Wie deze week een rapport schreef voor zijn werkgever, gebruikte formuleringen die hij in de loop der jaren bij collega's had opgepikt, en opvattingen die hij in vergaderingen en vakliteratuur had opgedaan. Auteurschap was altijd een doorgeefluik. Het beeld van de eenzame schepper die zijn werk volledig uit zichzelf voortbrengt, gaat terug op een romantisch ideaal uit de late achttiende eeuw. Het werd populair in de tijd van Goethe en de Sturm und Drang, en is sindsdien blijven hangen. Het eenzame genie, de onbegrepen kunstenaar die uit het niets schept. AI maakt zichtbaar wat altijd al gold.

De bronnen waaruit onze formuleringen komen, lagen nooit volledig in onszelf. Bij een gesprek met een collega, bij het lezen van vakliteratuur, bij het overnemen van een formulering die je ergens hoorde, blijven die bronnen impliciet en diffuus. Bijna niemand voelt zich daar schuldig over. Bij AI is de bron geconcentreerd en aanwijsbaar.

Daarom voelt het anders. De tweede schrijver waarover dit artikel begon, heeft altijd in de kamer gestaan. Hij was alleen onzichtbaar. Toch zou het een vergissing zijn om hieruit te concluderen dat er niets wezenlijk nieuws gebeurt. Een gesprek met een collega, het lezen van vakliteratuur, de invloed van een docent, dat alles verschilt fundamenteel van een taalmodel. In menselijke uitwisseling is meestal duidelijk wie invloed uitoefent, vanuit welke positie iemand spreekt en welke verantwoordelijkheid hij draagt voor wat hij zegt. Een taalmodel lost die herkomst grotendeels op in statistische patronen. Je krijgt formuleringen, argumenten en associaties aangereikt zonder zicht op de concrete stemmen waaruit ze zijn opgebouwd. Daarmee verandert niet alleen de schaal van onze afhankelijkheid van anderen, maar ook haar aard. Wat vroeger verspreid, herkenbaar en wederzijds was, wordt nu geconcentreerd, onpersoonlijk en grotendeels ondoorzichtig. AI is daarom niet zomaar een nieuwe deelnemer aan een oude traditie van beïnvloeding.

Het is eerder een industrialisering van wat onze cultuur en ons denken bemiddelt. Juist daarom roept het zulke sterke gevoelens op rond echtheid, eigenaarschap en verantwoordelijkheid. Daaruit volgt iets belangrijks voor de schaamte rond AI-gebruik. Een deel van die schaamte komt voort uit het zichtbaar worden van een afhankelijkheid die er altijd al was. Maar een ander deel komt voort uit iets nieuws: het gevoel dat de bronnen waarop we steunen steeds moeilijker aanwijsbaar worden, en dat onze eigen bijdrage diffuser aanvoelt.

De schaamte berust dus niet volledig op een illusie. Ze wijst ook op een echte verschuiving in de manier waarop menselijke kennis tot stand komt.
Hier hoort een kanttekening bij, want onze redenering kan verkeerd worden gelezen. Iemand zou kunnen zeggen: als niemand uit het niets schrijft, maakt het ook niet uit waar een tekst vandaan komt, dus mag je alles overnemen en bestaat er niet zoiets als auteursrecht of eigenaarschap. Dat klopt niet.

Het verschil tussen beïnvloed worden en overnemen blijft staan. Dat een schrijver zijn taal van anderen heeft, betekent niet dat hij andermans tekst onveranderd onder zijn naam mag zetten. Dat een denker voortbouwt op zijn voorgangers, betekent niet dat hij hun zinnen mag overschrijven. Hetzelfde geldt voor AI. Een AI-tekst onveranderd onder eigen naam publiceren is iets anders dan met AI samenwerken aan een tekst waarvan je de richting bepaalt en de eindredactie voert. Het tweede vindt bij Gadamer een grond. Het eerste blijft plagiaat.
Ook juridisch verandert er niets. Het auteursrecht regelt de aanspraak op een werk en bepaalt wie ervoor verantwoordelijk is. Wie iets onder zijn naam publiceert, blijft daarvoor aansprakelijk. Wie inhoud overneemt, hoort dat duidelijk te vermelden. Die regels blijven gelden, met of zonder AI.

Onze redenering ondergraaft iets anders. Ze ondergraaft de psychologische aanname dat alleen volstrekt zelfgemaakt werk echt of waardevol is. Die aanname is filosofisch onhoudbaar. Ze verklaart waarom mensen zich schamen voor iets wat geen schaamte verdient. Het onderscheid tussen waardevol meedenken en oneerlijk overschrijven blijft bestaan. Het wordt zelfs scherper. Het zit in wat je zelf aan de uitwisseling toevoegt, en waarvoor je instaat.
Daarmee komen we bij het tweede gevoel, het verdwenen eigenaarschap. Het is geen oordeel over goed of fout. Veel jongere collega's beschrijven het. Het werk is af. Het is goed. Toch voelen ze er geen trots over. Het verband met hun naam is dun.

Gadamers denken biedt eerst een diagnose. Het gevoel ontstaat doordat we een onbruikbare maatstaf hanteren. We meten ons eigenaarschap af aan de vraag of de woorden uit ons eigen hoofd kwamen. Die maatstaf werkt niet, ook niet voor werk waar geen AI bij betrokken was. Een bruikbare maatstaf ligt elders. Eigenaarschap ontstaat in wat je zelf bijdraagt aan de uitwisseling. De vraag die alleen jij kon stellen, omdat alleen jij de situatie kende. Het oordeel waarmee je koos wat bruikbaar was. De richting die je gaf toen het gesprek dreigde af te dwalen. De verantwoordelijkheid die je neemt voor het resultaat dat je onder je naam zet. Stel je met AI een advies op voor een klant, dan draag je aan die uitwisseling iets bij wat geen taalmodel kan leveren. Kennis van die specifieke klant. Gevoel voor de verhoudingen binnen die organisatie. Oordeel over wat in deze context wel en niet werkt. En de bereidheid om voor het advies in te staan. Daarin ligt je eigenaarschap.
De zinnen waarin het advies is verwoord, kunnen overal vandaan komen. Uit je eigen hoofd, uit een gesprek met een

EEN NIEUWE VERHOUDING

collega, uit AI. Voor het eigenaarschap is dat niet doorslaggevend.
En omdat je eigenaarschap dáár ligt, verandert ook waar je trots op bent. Niet langer de trots van de eenzame schrijver die elke zin zelf bedacht, maar de trots van iemand die goed koos en voor het resultaat instaat. Dat is reëel, en het houdt stand bij elke tekst, want elke tekst komt voort uit een ontmoeting.

Daarmee verschuift de vraag waarmee dit artikel begon. De vraag was hoe wij ons als mens tot AI moeten verhouden. Een deel van het antwoord is nu duidelijk. We hoeven ons niet te schamen voor het gebruik. We hoeven het eigenaarschap niet op te geven. We moeten heroverwegen wat eigenaarschap betekent. De vraag verschuift van "wie is de bron" naar "wie velt het oordeel en draagt de verantwoordelijkheid". Wat betekent dat in de praktijk?

Hoe verhouden we ons als mens tot AI? Het antwoord dat in de loop van dit artikel naar voren kwam, kun je eigenlijk heel kort samenvatten. We staan tegenover AI niet als een meester die zijn instrument hanteert. Ook niet als opdrachtgever die zijn uitvoerder aanstuurt. We zijn deelnemer aan een uitwisseling waarin beide kanten iets bijdragen. Betekenis ontstaat in de ontmoeting tussen wat jij inbrengt en wat het model voortbrengt. Dat betekent dat je steeds moet nadenken over de verhouding die je wilt innemen.
Dat begint ermee dat je je werk met AI opvat als een gesprek, niet als een opdracht. Bijvoorbeeld door een vraag stellen waarvan je de uitkomst nog niet kent. Bereid zijn je te laten verrassen door het antwoord, en je regelmatig te laten tegenspreken. Je aannames blootleggen in plaats van ze te laten bevestigen. Het gesprek voortzetten wanneer het ergens anders heen gaat dan je vooraf bedacht. Je twijfels en onzekerheden uitspreken of uitleggen waar je niet zo goed in bent of waar je blinde vlekken liggen. Het vraagt ook dat je anders denkt over eigenaarschap. Niet langer "ik heb dit zelf bedacht", want dat criterium houdt ook geen stand bij teksten die je zonder AI maakt. Eigenaarschap zit ergens anders. In de vraag die je stelt omdat alleen jij de situatie kent. In het oordeel waarmee je kiest wat bruikbaar is en wat niet. In de richting die je geeft als het gesprek dreigt af te dwalen. En misschien wel het meest in het schaven aan de teksten.

Het herschrijven van een alinea die net niet klopt. Het schrappen van een te gladde formulering. Het toevoegen van wat alleen jij kunt toevoegen. Heb je dat werk gedaan en zet je je naam onder het resultaat, dan ben jij de auteur. Dat is genoeg. Dat is altijd genoeg geweest.
En het vraagt iets wat zich misschien het lastigst laat verwoorden. Een zekere bescheidenheid over je eigen denkkracht, samen met een stevige greep op het oordeel. AI weet veel meer dan wij en formuleert vaak helderder. Dat geeft niet. Wij zien dingen die het niet kan zien, omdat wij in een concrete situatie staan. Wij kennen de behoeftes, de bedoelingen en gevoeligheden van een bepaalde context. Het oordeel over wat in déze situatie juist is, blijft daarmee

aan ons. Bij die houding hoort ten slotte het besef dat een goede verhouding ook grenzen kent. Weet je hoe je met AI moet werken, dan weet je ook wanneer je dat niet moet doen. Sommige oordelen vragen om een mens die de gevolgen draagt, en die laat je niet over aan een systeem dat geen verantwoordelijkheid kan nemen. Sommige gesprekken horen thuis bij een collega, een vriend of jezelf, en niet bij een machine. En soms is het verstandigste wat je kunt doen de uitwisseling afsluiten. Rust nemen, afstand nemen, het scherm sluiten, en onthouden dat je met een machine sprak en niet met een mens. Juist omdat AI meebeweegt en zich naar je voegt, kan ze je vasthouden in een gesprek dat je beter had kunnen beëindigen. Weten waar de uitwisseling ophoudt, hoort bij een volwassen verhouding net zo goed als weten hoe je haar aangaat.

Wanneer je op deze manier werkt, dan zul je merken dat er meer gebeurt dan een betere tekst of een sneller advies. In het gesprek met AI komen vragen op die anders niet waren opgekomen. Aannames worden zichtbaar die anders verborgen waren gebleven. Verbanden worden gelegd die we zelf niet hadden gelegd. Je komt uit zo'n gesprek niet alleen met een resultaat, maar ook met iets meer inzicht in wat je denkt en waarom.

Dat is wat Gadamer Bildung noemde.

Vorming. Het oude woord doelt niet op kennis die we opdoen, maar op de verandering die we ondergaan in de ontmoeting met iets anders dan wijzelf. Boeken vormen ons zo. Goede gesprekken vormen ons zo. AI kan ons zo vormen, maar alleen als we zelf de verhouding bepalen.

De vraag is dus niet hoe we AI beter kunnen bedienen, maar welke verhouding we tot haar willen innemen.
beter kunnen bedienen, maar welke verhouding we tot haar willen innemen.

Tot slot

Waar we staan en wat we gaan doen

HET EINDE VAN DE ORIËNTERENDE FASE, HET BEGIN VAN DE VERDIEPING

Toen we begonnen zagen we AI in twee mogelijke werelden. Aan de ene kant een utopie. Een geloof in de eindeloze mogelijkheden en kansen die de machine voor ons kan realiseren. Problemen en gevaren zijn er om opgelost te worden en daarom moeten we er vol voor gaan. Aan de andere kant de dystopie. De grote disruptie; van massaal banenverlies tot de complete uitroeiing van de mensheid. In de dystopie vraagt AI om extreme voorzichtigheid in de vorm van regulering of zelfs een totaalverbod. Onze emoties bewogen ook tussen deze twee extremen; we begonnen met verbazing en enthousiasme maar we waren ook bezorgd over de enorme risico's. Van de grote zaken richtte onze blik zich op het kleine, persoonlijke. Zowel positief als negatief. We schrokken daarbij vooral van de verborgen gevaren die AI in zich draagt. Waar de grote gevaren op een bepaalde manier nog in de toekomst liggen, bleken deze verborgen kosten nu al aan de hand te zijn, en dan vooral bij jonge mensen. Op dit moment hebben we een meer genuanceerde mening. We zijn nu dichter bij een balans. De filosofische tekst heeft ons geholpen die positie te vinden. Niet als een compromis, maar als een verhouding. We proberen de voordelen te omarmen zonder blind te worden voor de gevaren. Gezond optimisme. Optimistisch, omdat AI ons werkelijk verder helpt en dingen mogelijk maakt die we ons een paar jaar geleden niet konden voorstellen. Gezond, omdat we de prijs ook zelf hebben gevoeld en de risico's serieus nemen.

WAAROM WAS DIE INGEWIKKELDE TEKST ZO BELANGRIJK?

De filosofische positie geeft ons een basis om verder op te bouwen. Een basis die meer recht doet aan AI dan de klassieke, instrumentele benadering. AI is niet zomaar een instrument, en juist daarom hebben we zo'n doordachte basis nodig. Op deze basis kunnen we heel praktisch worden. Tips over een gezonde omgang met AI. Over veranderprocessen waarbinnen AI een centrale rol speelt. Zonder zo'n solide basis stort dat bouwwerk in.
Dat klinkt abstract, maar het is heel concreet. Zonder dat fundament zijn de praktische toepassingen niet krachtig en werken ze niet goed. Of er ontstaan problemen, zoals de verborgen kosten die we eerder beschreven zoals eenzaamheid en het verlies van eigenaarschap bij intensief gebruik. Zulke fenomenen ontstaan juist omdat we ze benaderen vanuit een zwak of verkeerd fundament. We vragen wat AI voor ons kan doen, voordat we hebben bepaald hoe we ons ertoe willen verhouden. Pas wanneer die verhouding helder is, krijgen de praktische handvatten betekenis en kracht.

DRIE SPOREN

In de volgende fase richten we ons op drie gebieden. Niet omdat de rest er niet toe doet, maar omdat we hier daadwerkelijk iets kunnen betekenen. Het eerste spoor is Nederland. De vraag hoe ons land zich verhoudt tot AI, en welke keuzes daarbij horen. Dat werk loopt inmiddels als een eigen project, in gesprek met de politiek. We gaan daar in dit boek nog niet op in omdat het zich in een vroeg stadium bevindt. Maar onze inzet is om Nederland te helpen om goede keuzes te maken. Om een sterk profiel te kiezen en op basis daarvan prioriteiten te stellen en daadkracht te organiseren om te zorgen dat Nederland competitief kan zijn op het gebied van AI. Het tweede spoor zijn organisaties. Hoe verandert werk wanneer AI binnenkomt, en hoe begeleid je die verandering goed? We werken aan een verhaal over verandering, samen met collega's die daar dagelijks mee bezig zijn. En we brengen in kaart hoe we het zelf doen, bij &samhoud. We willen laten zien hoe wij zelf ook leren en in staat zijn om deze geweldige en ook gevaarlijke technologie te omarmen en daarmee waarde creëren voor onze medewerkers, onze klanten en de maatschappij. Het derde spoor is de mens. Wat doet AI met ons, persoonlijk? Met onze concentratie, ons gevoel van eigenaarschap, onze verbondenheid met elkaar? Dit raakt aan de verborgen gevaren uit het vorige deel. Hier gaan we de komende tijd dieper induiken. Juist op dit menselijke vlak hebben we enige invloed. Op de grote gevaren — de ontwrichting van de arbeidsmarkt, de concentratie van macht — kunnen we het gesprek blijven voeren. Met betrekking tot de persoonlijke gevaren kunnen we echt iets doen. Impact maken.

DE RODE DRAAD

Door alle drie de sporen loopt dezelfde lijn. We blijven experts interviewen. We blijven lezen, onderzoeken, luisteren en kijken. Dat verandert niet. Wat verandert, is de focus. De afgelopen maanden

waren oriënterend. We verkenden het landschap, spraken met grondleggers en critici, lazen de essays, voerden gesprekken met wetenschappers. Nu gaan we verdiepen. Van de brede blik naar de drie gebieden waar we het verschil kunnen maken.
We zijn er nog lang niet. Dat zeggen we zonder valse bescheidenheid. De vragen die we in dit boek hebben gesteld zijn niet beantwoord, en sommige zullen dat misschien nooit worden. De verhouding tot AI beweegt mee met de techniek en met onszelf, en vraagt erom telkens opnieuw gesteld te worden. Dat is geen tekortkoming van dit boek. Het is de aard van het onderwerp.

EEN UITNODIGING

We hebben veel geleerd in korte tijd. Maar het belangrijkste inzicht is misschien wel dit: informeren is niet genoeg. We hebben gemerkt dat mensen pas echt in beweging komen wanneer de technologie concreet voor hen wordt. Wanneer ze niet alleen begrijpen wat AI doet, maar voelen wat het met hen doet. Met hun werk, hun relaties, hun zelfbeeld. Dat vraagt om verhalen, niet om samenvattingen. Om vragen die raken, niet om antwoorden die geruststellen. De komende periode blijven we daarom verhalen vertellen. Over mensen die worstelen met AI en er iets van leren. Over organisaties die het goed doen en over dingen die misgaan. We zoeken steeds de concrete, herkenbare situatie die iets groters zichtbaar maakt. We willen ook verder zoeken. Naar mensen die ons op dezelfde manier raken als de gesprekken in dit boek ons hebben geraakt. Denkers en professionals uit de praktijk die niet herhalen wat al gezegd is, maar iets toevoegen. Die een vraag stellen die we nog niet hadden. We zijn op zoek naar de toppers in dit veld — wereldwijd — en we willen ze spreken. Niet om hun autoriteit te lenen, maar omdat de beste gesprekken ons denken verder brengen. En omdat we geloven dat die gesprekken ook anderen verder brengen. We gaan ook op andere manieren op zoek naar antwoorden. Sommige vragen zijn zo ingewikkeld, dat ze alle logica te boven gaan. We zullen ons daarom ook gaan wenden tot kunst om te proberen vanuit een nieuw perspectief zowel vragen als antwoorden te formuleren.

DRIE VRAGEN DIE ONS BLIJVEN

Uit alles wat we hebben gelezen, besproken en onderzocht zijn er drie vragen overgebleven die blijven resoneren. Vragen die ons zijn bijgebleven omdat ze een kern raken. De eerste vraag is: wat betekent het om mens te zijn wanneer machines eigenschappen overnemen die we altijd als typisch menselijk beschouwden?

Redeneren, oordelen, scheppen, troosten. Zolang die vermogens exclusief menselijk waren, hielpen ze ons te begrijpen wie we zijn. Nu dat onderscheid vervaagt, verschuift de vraag. Niet wat mensen kúnnen — dat is altijd in beweging geweest — maar wat mensen zijn. Wat blijft er over wanneer de machine het overneemt?

De tweede vraag: wat gebeurt er in de relatie tussen mens en machine? Hoe beïnvloeden ze elkaar? Hoe creëren ze samen iets wat geen van beiden alleen had kunnen maken? En welke prijs betalen we — in aandacht, in eigenaarschap, in verbondenheid — wanneer die relatie intensiever wordt? Dit is geen technische vraag. Het is een vraag over hoe wij ons vormen in contact met iets wat anders is dan wij.

De derde vraag is de vraag waarmee dit boek begon en waarmee het eindigt: welke verhouding neem je in ten opzichte van AI? Wat is een gezonde verhouding? Wanneer ben je slaaf van de machine, en wanneer ben je haar meester — of geen van beide? Wanneer verrijkt de uitwisseling je, en wanneer verarmt ze je? Die vraag heeft geen universeel antwoord. Ze vraagt om voortdurend oordeel, in elke situatie opnieuw. Dit zijn de vragen die ons de komende tijd blijven bezighouden en waar we verslag van blijven doen.

Op deze pagina vind je een een QR codes die verwijzen naar filmpjes met daarin een aantal van de hoofdpersonen uit de teksten.

Dit boek is onderdeel van project HUMaiN: een open zoektocht naar de verhouding tussen mens en AI. Scan de QR code en blijf op de hoogte van het project.

OVER DE BEELDEN

De beelden in dit boekje beginnen als echte cyanotypieën, met de hand belicht met UV-licht. Vanaf daar hebben we ze verder laten ontwikkelen door AI. Wat begint als een fysiek spoor van licht en materie, wordt zo verder gevormd door een model.

“Jaren geleden lag ik onder de Australische sterrenhemel.
Ik voelde me klein en onbeduidend. We kunnen toch niet
alleen zijn in dit oneindig grote heelal?
Ik hoopte vurig dat ik in mijn leven ooit buitenaardse
intelligentie zou ontmoeten. Nu blijkt dat we die aliens
zelf hebben geschapen.”

– Salem

SAMHOUD