

Ethan Mollick

CO-INTELLIGENTIE

Slimmer werken met AI

2026-EDITIE

Vertaald door
Jan Willem Reitsma

POM
P R E S S

Eerste tot en met negende druk, september 2024-januari 2025
Tiende tot en met veertiende druk, 2025-editie, april 2025-juli 2025
Vijftiende druk, 2026-editie, november 2026

Oorspronkelijke titel *Co-intelligence. Living and working with AI*
Copyright © 2024 Ethan Mollick
All rights reserved.

Nederlandse vertaling
Copyright © 2024 POM.press (een imprint van Maven Publishing B.V., Amsterdam) / Jan Willem Reitsma

2026-editie
Copyright © 2025 Ethan Mollick

Nederlandse vertaling
Copyright © 2025 POM.press / Jan Willem Reitsma

www.pom.press

Ontwerp omslag Valeria Pieuniou / DPS
Vormgeving binnenwerk Elgraphic
Drukwerk Wilco

ISBN 978 94 9343 427 1
ISBN e-boek 978 94 9343 428 8
NUR 801

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm, kwantum-entanglement met een superintelligent AI-bewustzijn dat zich uitstrekt over meerdere universa of op welke wijze en/of door welk ander medium ook, zonder voorafgaande schriftelijke toestemming van de uitgever. Niet dat dat allemaal veel uitmaakt als de AGI er straks is. Maar toch.

INHOUD

Voorwoord door Alexander Klöpping	9
Inleiding: Drie slapeloze nachten	13

DEEL 1

1 De geesten van aliens	22
2 Op één lijn met de alien	42
3 Vier principes voor co-intelligentie	57

DEEL 2

4 AI als mens	72
5 AI als creatieve denker	96
6 AI als collega	121
7 AI als privédocent	151
8 AI als coach	167
9 AI als onze toekomst	179

Epiloog: AI als wij	195
----------------------------	-----

2026-EDITIE

1	Aan de slag met AI: goed genoeg prompts	198
2	Vijftien keer AI wel gebruiken, en vijf keer niet	205
3	Het eind van de zoektocht, het begin van het onderzoek	211
4	AI voor je laten werken: Leiderschap, het Lab en de Massa	218
5	Tegen 'hersensletsel'	228
6	GPT-5: het doet gewoon dingen	238
7	Massa-intelligentie	246
8	Samenwerken met tovenaars	255
	Woord van dank	263
	Noten	265

VOORWOORD

door Alexander Klöpping

Sinds de dotcombubbel heb ik geen grotere hype meegemaakt dan het huidige rumoer rondom kunstmatige intelligentie. Silicon Valley smijt met geld, bedrijven vragen zich in paniek af of ze wel ‘*AI ready*’ zijn en overheden buitelen over elkaar heen om regelgeving te bedingen.

Er is ook écht wat aan de hand. De snelheid van de technologische vooruitgang is op dit moment ongekend. Iedereen, en ikzelf ook, vraagt zich af: wat betekent dit voor mij? 9

Om weer grip te krijgen en te doorgronden hoe AI je werk en leven gaat veranderen in de komende twee jaar, moet je vandaag beginnen met het dagelijks gebruiken van AI. Je zult versteld staan hoe nuttig, leuk of vermakelijk het kan zijn. Ethan Mollick neemt je in dit boek bij de hand.

Alleen als je AI vanaf nu dagelijks gaat gebruiken, ontwikkel je een gevoel voor de snelheid van de vooruitgang. Soms werkt AI nog niet perfect. Maar dingen die vandaag nog niet goed werken, doen dat morgen misschien wel. Als je de ontwikkelingssnelheid doorvoelt, levert dat, volgens Ethan Mollick in zijn eerste hoofdstuk, drie slapeloze nachten op. Maar daarna begint het avontuur pas echt.

De meeste Nederlanders gebruiken nog bijna geen AI. Maar een beroepsgroep die dat al wel veel doet zijn programmeurs. Als ik bevriende programmeurs vraag naar de impact van AI op hun werk, dan hoor ik stevast hetzelfde: dezelfde klus kost nu de helft minder tijd dan

twee jaar geleden, omdat ze sámen met AI programmeren. Laat dat even bezinken. Ze doen opeens hetzelfde werk, in de helft van de tijd.

Deze trend zal zich doorzetten naar andere vakgebieden. Advocaten, journalisten, marketingmensen, accountants – iedereen die met informatie werkt (en dikke kans dat als je dit boek leest jij ook veel met informatie werkt), krijgt te maken met dezelfde soort verandering. Nu bedien je een computer nog met een muis en toetsen, maar straks krijg je onzichtbare, persoonlijke assistenten die kunnen anticiperen op wat je wilt doen en veel van je werk alvast uitvoeren.

Het is een patroon dat zich nog veel zal herhalen: eerst komen nieuwe technologische mogelijkheden en daarna volgen de praktische toepassingen.

Technisch is het nu bijvoorbeeld mogelijk om een video waarin iemand een bepaalde taal spreekt, met een AI-tool te vertalen naar een andere taal. De stem wordt gekloond, de tekst wordt in een andere taal uitgesproken, en de mondbeweging van de video wordt aangepast (kijk voor een voorbeeld op pom.press/heygen).

De techniek is er. Vervolgens is de vraag: hoe snel past onze cultuur zich aan aan de mogelijkheden? Hoelang duurt het bijvoorbeeld nog voordat buitenlandse wereldleiders voortaan Nederlands spreken in het NOS Journaal? Wanneer luister je met hetzelfde gemak naar een Japanse podcast (door AI vertaald) als een Nederlandse?

Ethan Mollick stelt in dit boek dat je moet beseffen dat de AI die je vandaag gebruikt, de slechtste AI is die je ooit zult gebruiken. Ik probeer dat in m'n oren te knopen en probeer me regelmatig voor te stellen hoe de wereld er over een paar jaar uitziet. Ik doe voorspellingen door de lijntjes van de technologische ontwikkelingen door te trekken. Zo denk ik dat alle informatie die je binnenkort tot je neemt, op een of andere manier door AI is bewerkt. Als je een uur durende podcast te lang vindt, kan AI die voor je inkorten tot een half uur. Of je vindt een podcast zo leuk dat je wilt dat hij dubbel zo lang duurt. Geen probleem! Laat AI de rest verzinnen met diepgaande, relevante en gefactcheckte content. Hetzelfde zal gelden voor nieuws, boeken, films en hoorcolleges.

Voorwoord

Nu wordt de keuze van welke content we consumeren vaak gemaakt door AI. Denk aan de algoritmes van Spotify of YouTube die bepalen welke nummers of video's we te zien krijgen. Maar binnenkort zal AI een groot deel van onze cultuur genereren. Ik heb voorbeelden gehoord van muziek gemaakt door de AI-dienst Suno, die diepe Southern Blues speelt met een stem doorspekt van pijn en een gitaar waar de emotie vanaf spat (luister zelf via pom.press/suno). Het is een geest in een machine in plaats van een echt mens, maar de emotie is er niet minder om.

Kortom: de snelle technologische ontwikkelingen en toegankelijkheid gaan impact hebben op onze cultuur. Ik denk dat we snel zullen wennen aan allerlei vormen van AI. Bij de overheid, het bedrijfsleven, het onderwijs, de zorg – hoe we elkaar ontmoeten, waar we over praten, wat we lezen. Het zal onze cultuur gaan bepalen. Ons werk. Ons leven.

Er is genoeg om optimistisch over te zijn: het wordt opeens voorstelbaar dat elke leerling in Nederland het beste één-op-één onderwijs kan krijgen van AI-docenten. Er zijn talloze start-ups die proberen AI-modellen te gebruiken om nieuwe medicijnen uit te vinden of nieuwe grondstoffen te ontwikkelen. Misschien gaat AI een cruciale rol spelen in de manier waarop we de klimaat- en energiecrisis kunnen bestrijden.

11

Maar het is ook makkelijk om pessimistisch te worden: leven we straks nog meer in filterbubbels met onbetrouwbare informatie? Wat kunnen criminelen of vijandige overheden met deze tools? Wat als de macht van techbedrijven nog groter wordt? Wat als ze nóg meer energie gaan gebruiken bij het trainen van AI?

Dit alles brengt me in grote verwarring. Het is gemakkelijk om overweldigd te raken door de snelheid en omvang van deze veranderingen. Ethan Mollick geeft ons de komende pagina's meer grip. Maar, zoals gezegd, we beginnen met drie slapeloze nachten.

PS Dit stuk is geschreven door AI nadat ik een warrig verhaal heb ingesproken in een microfoon en Claude 3.5 Sonnet (een concurrent van ChatGPT) de tekst heeft bewerkt. Eerst heb ik een aantal stuk-

ken tekst ingevoerd die ik eerder zelf schreef, en aan Claude gevraagd om mijn schrijfstijl te beschrijven. Vervolgens heb ik die output gebruikt voor de volgende prompt:

We gaan een opiniestuk schrijven met de volgende richtlijnen:

1. Wees direct en to-the-point. Kom snel tot de kern van je argument zonder onnodige omhaal.
2. Neem een duidelijke stelling in. Wees niet bang om je mening te geven over het onderwerp.
3. Wees kritisch.
4. Gebruik beeldende taal. Maak gebruik van levendige voorbeelden en metaforen.
6. Bouw je betoog op met duidelijke argumenten en voorbeelden.
7. Voeg een subtiele ondertoon van humor en ironie toe.
8. Schroom niet om persoonlijke ervaringen of observaties te delen.
- 12 9. Schrijf prikkelend, zet de lezer aan het denken en probeer.
10. Balanceer je meningen met feitelijke informatie en context.
11. Houd je taal toegankelijk en begrijpelijk zonder jargon.
12. Eindig met een krachtige, samenvattende zin.

Pas deze stijlelementen toe op een manier die past bij het onderwerp en de toon van het stuk. Het doel is om de lezer te engageren, informeren en aan het denken te zetten, zoals Alexander Klöpping.

Voeg nu een stuk tekst toe. Dit is een transcript van wat ik in de microfoon heb ingesproken. Het is ongepolijst en 'stream of consciousness'. Maak er een goedlopende tekst van. Haal dubbelingen weg. Zet alles in logischere volgorde, zodat het een verhaal wordt met een logische opbouw. Maak een samenhangend verhaal. Pas de inhoud niet aan, maar herschrijf naar mijn schrijfstijl.

Dit voorwoord is het resultaat. Ik moest het een beetje bijschaven. Nog wel.

INLEIDING

Drie slapeloze nachten

Volgens mij is de prijs die je moet betalen om AI te leren kennen – om AI écht te leren kennen – ten minste drie slapeloze nachten.

Na een paar uur gebruikmaken van generatieve AI-systemen komt er een moment dat je inziet dat *Large Language Models* (LLM's), de nieuwe vorm van AI achter diensten zoals ChatGPT, zich anders gedragen dan je van een computer zou verwachten. Ze gedragen zich eerder als een mens. Het begint tot je doorte dringen dat je met iets nieuws, iets heel vreemdsoortigs te maken hebt, en dat alles op het punt staat te veranderen. Je blijft wakker, opgewonden en zenuwachtig tegelijk, en vraagt je af: hoe zal mijn baan er later uitzien? Wat voor baan zullen mijn kinderen kunnen krijgen? Kan dit ding nadenken? Midden in de nacht loop je terug naar je computer en doe je schijnbaar onmogelijke verzoeken, alleen maar om te zien of AI daaraan kan voldoen. Je beseft dat de wereld fundamenteel is veranderd en dat niemand je goed kan zeggen hoe de toekomst eruit zal zien.

13

Ik ben weliswaar geen informaticus, maar wel een academicus die innovatie bestudeert en lang betrokken is geweest bij onderzoek naar toepassingen van AI, vooral in het onderwijs. Door de jaren heen heeft AI veel meer beloofd dan ze heeft geleverd.* Decennialang

* Noot van de vertaler: Het woordgeslacht van AI en ChatGPT levert in het Nederlands een dilemma op. Ik volg hierbij de regels van het Groene boekje en de Taalunie. AI is vrouwelijk omdat intelligentie vrouwelijk is, ChatGPT is mannelijk omdat zowel (ro)bot als transformer mannelijk zijn. (ChatGPT antwoordt in beide gevallen dat ze onzijdig zijn.)

leek AI-onderzoek steeds op het punt van een grote doorbraak te staan, maar de meeste praktische toepassingen, van zelfrijdende auto's tot geïndividualiseerd onderwijs, vorderden maar met een slak-kengang.¹ Al die tijd bleef ik experimenteren met tools als de GPT-modellen van OpenAI om te kijken hoe ik die in mijn werk kon inpassen, en tijdens mijn colleges gaf ik studenten opdrachten waarbij ze AI moesten gebruiken. Met andere woorden: mijn slapeloze nachten kwamen al vroeg, kort na de lancering van ChatGPT in november 2022.

14 Al na een paar uur was het duidelijk dat er een enorm verschil zat tussen de eerdere edities van GPT en deze nieuwe versie. Vier dagen na de lancering van ChatGPT besloot ik dit nieuwe stuk gereedschap te demonstreren tijdens mijn college Ondernemerschap voor bachelorstudenten. Vrijwel niemand had ervan gehoord. Ik gaf een demonstratie voor mijn studenten om hun te laten zien hoe AI kan helpen om ideeën te generen, bedrijfsplannen te schrijven, die bedrijfsplannen in gedichten te veranderen (niet dat daar veel vraag naar is), en in algemene zin de rol van een startende ondernemer te vervullen. Aan het eind van het college had een van mijn studenten, Kirill Naumov, een werkende demo gemaakt voor zijn werkgroep Ondernemerschap: een op Harry Potter geïnspireerde bewegende fotolijst die reageerde als mensen langsliepen, met behulp van software die hij nooit eerder had gebruikt, in nog niet de helft van de tijd die dat anders zou hebben gekost. Aan het eind van de volgende dag waren geïnteresseerde investeerders al naar hem op zoek.

Nog geen twee dagen nadat ik mijn studenten had laten kennismaken met AI, vertelden meerderen van hen dat ze ChatGPT hadden gebruikt om begrippen die ze verwarrend vonden aan hen te laten uitleggen 'alsof ze tien jaar oud waren'. Ze staken niet langer hun hand op, want waarom zou je jezelf tijdens college voor schut zetten als je het later aan de AI kon vragen? En elk essay was opeens volmaakt grammaticaal geschreven (hoewel referenties vaak niet klopten en de laatste alinea doorgaans met 'Conclusie' begon – een dui-

delijk teken van ChatGPT-gebruik uit de begintijd, dat later is verholpen). De studenten waren echter niet alleen enthousiast, maar ook zenuwachtig. Ze wilden weten hoe de toekomst eruitzag.

Sommigen van hen vroegen wat de gevolgen voor hun toekomstige carrière zouden zijn ('Moet ik radioloog worden als AI een groot deel van mijn werk kan doen?' 'Is het schrijven van reclameteksten over vijf jaar nog een goede baan?'). Anderen vroegen wanneer deze technologische ontwikkeling zou eindigen, en of die wel zou eindigen. Iemand vroeg zelfs of *Artificial General Intelligence* (AGI), de hypothetische machine die slimmer dan een mens is, nog voor hun afstuderen gerealiseerd zou worden.

Ik had daar toen geen antwoord op (hoewel ik inmiddels iets meer antwoorden heb), maar besepte wel dat ik niet immuun was voor die veranderingen. Daarom besloot ik zelf aan de slag te gaan en de AI op de proef te stellen. In de afgelopen vijf jaar had ik me vooral beziggehouden met het bouwen van simulaties om *business skills* bij te brengen. Er was een forse investering van mijn universiteit en de inzet van een tiental talentvolle mensen voor nodig om die gedetailleerde digitale vaardigheden te bouwen. Na duizenden uren werk waren het geweldige games geworden: met duizenden coderegels konden we software voor leersimulaties realiseren die hielpen bij het bijbrengen van vaardigheden zoals onderhandelen. Maar ik besloot iets bij ChatGPT in te tikken:

15

Je wordt mijn docent Onderhandelen. Je simuleert een gedetailleerd scenario waarin ik moet onderhandelen. Jij speelt de rol van de ene partij en ik speel die van de andere. Bij elke stap van het scenario vraag je me wat mijn antwoord is en wacht je tot je dat krijgt. Nadat je mijn antwoord hebt gekregen, vertel je me wat de andere partij doet en zegt. Je waardeert mijn antwoord met een cijfer en geeft gedetailleerde feedback over wat ik beter kan doen met behulp van wetenschappelijke kennis van onderhandelen. Als ik goed presteer, geef je me een moeilijker scenario, en als ik slecht presteer een makkelijker scenario.

ChatGPT antwoordde:

■ Prima, ik help je graag bij het oefenen in onderhandelen met een eenvoudige simulatie-oefening! Laten we met een eenvoudig scenario beginnen. Je bent een verkoper die 100 pennen aan een klant probeert te verkopen. De pennen kosten normaal gesproken één dollar, maar de klant probeert af te dingen. De klant opent met het aanbod om de pennen voor vijftig dollarcent per stuk te kopen. Wat is daarop jouw antwoord?

En vervolgens leidde hij me door een complete simulatie, die weliswaar niet perfect was, maar 80 procent deed van wat ons team maanden werk zou hebben gekost. En dat na een *prompt* van slechts één alinea!

Zo begonnen mijn slapeloze nachten.

16 Eerlijk gezegd zijn die nog steeds niet voorbij. Er gaat geen week voorbij zonder dat AI ons versteld doet staan – soms door een bewonderenswaardige en soms door een zorgwekkende ontwikkeling. Ik raakte al gauw verstrikt in een ruzie met een chatbot die me van onethisch gedrag betichtte toen ik hem vroeg mij te helpen bij het schrijven van software. Ik ontdekte dat ik mooie plaatjes kon maken door een machine de juiste woorden te voeren. Ik kon software schrijven in Python, een programmeertaal die ik nooit had geleerd. Ik ontdekte dat een machine een groot deel van mijn werk kon doen... en misschien wel dat van een heleboel andere mensen. Ik ontdekte iets wat opmerkelijk veel leek op een buitenaardse co-intelligentie, een die goed met mensen kon omgaan, zonder zelf menselijk of zelfs maar bewust te zijn. Volgens mij hebben we binnenkort allemaal onze drie slapeloze nachten.

En met mijn huidige slaapttekort keer ik telkens terug naar de vragen van mijn studenten: wat zal deze technologie betekenen voor de toekomst van de arbeid en het onderwijs? De ontwikkelingen volgen elkaar zo snel op dat het moeilijk te zeggen is, maar de eerste contouren tekenen zich af.

AI is iets wat technologieonderzoekers zoals ik een *General Purpose Technology* (ironisch genoeg ook afgekort tot GPT) noemen. Dit zijn zeldzame innovaties die eens per generatie optreden, zoals stoomaandrijving of internet, en die elke bedrijfstak en alle aspecten van het leven raken. In sommige opzichten is generatieve AI misschien zelfs nog groter.

General Purpose Technologies vinden meestal maar langzaam ingang, omdat ze veel andere technologieën nodig hebben om goed te kunnen functioneren. Internet is een uitstekend voorbeeld. Dat ontstond eind jaren 1960 als ARPANET, maar het duurde bijna drie decennia, tot in de jaren 1990, voordat het algemeen in gebruik kwam dankzij de uitvinding van de internetbrowser, de ontwikkeling van betaalbare computers en de groeiende infrastructuur die snel internet mogelijk maakte. Het duurde vijftig jaar voordat smartphones de opkomst van de sociale media mogelijk maakten. Veel bedrijven hebben internet nog steeds niet helemaal geaccepteerd: een bedrijf 'digitaal' maken is nog steeds onderwerp van felle discussies op economische universiteiten, vooral omdat veel banken nog altijd mainframecomputers gebruiken. Bij eerdere General Purpose Technologies duurde het vanaf hun ontstaan ook vele decennia tot ze succes behaalden. Neem computers, ook al zo'n transformerende technologie. De eerste computers werden in hoog tempo beter, dankzij de Wet van Moore, de reeds lang bestaande trend dat het rekenvermogen van computers zich elke twee jaar verdubbelt.² Toch duurde het tientallen jaren voordat computers in bedrijven en op scholen opdoken omdat ze, zelfs met hun razendsnel toenemende vermogen, vanaf een zeer primitief uitgangspunt waren begonnen. Maar Large Language Models bleken al enkele jaren nadat ze waren uitgevonden ongelooflijk capabel te zijn. Ze werden ook heel snel door consumenten geaccepteerd: ChatGPT had sneller 100 miljoen gebruikers dan welk product dan ook uit de geschiedenis, daarbij geholpen door het feit dat het gratis toegankelijk en ook voor privégebruik was, en ongelooflijk nuttig.³

Large Language Models worden ook beter. Hun omvang groeit

jaarlijks met een factor twee of nog meer, dus hun vaardigheden nemen ook toe. Ook al zal die vooruitgang waarschijnlijk gaan vertragen, ze voltrekt zich in een tempo dat alle andere grote technologieën in de schaduw stelt, en dan zijn LLM's nog maar een van de vele mogelijke *machine learning*-technologieën die de nieuwe AI-golf voortstuwen. Zelfs als de ontwikkeling van AI tot stilstand kwam terwijl ik deze zin afrond, zou ze ons leven nog steeds transformeren.

18 Ten slotte, hoe geweldig eerdere General Purpose Technologies ook waren, hun impact op de arbeid en het onderwijs was misschien geringer dan die van AI. Terwijl eerdere technologische revoluties vaak waren gericht op mechanische en repetitieve arbeid, werkt AI in veel opzichten als een co-intelligentie. Ze breidt het menselijk denken uit en kan het in potentie vervangen, met dramatische gevolgen. Uit vroege onderzoeken naar de effecten van AI bleek dat ze vaak voor een verbetering van 20 tot 80 procent in de productiviteit kan zorgen in sterk uiteenlopende werkzaamheden, van software schrijven tot marketing. Toen de stoomaandrijving, die meest elementaire vorm van General Purpose Technology die de Industriële Revolutie in gang zette, in een fabriek werd toegepast, steeg de productiviteit met maar 18 tot 22 procent.⁴ En ondanks decennialange observaties hadden economen de afgelopen twintig jaar moeite met het aantonen van een werkelijk langetermijneffect van computers en internet op de productiviteit.⁵

Bovendien hebben General Purpose Technologies niet alleen met werk te maken: ze raken aan elk aspect van ons leven. Ze veranderen hoe we lesgeven, hoe we ons vermaken, hoe we met andere mensen omgaan en zelfs hoe we naar onszelf kijken. Op basis van de eerste generatie AI's schreeuwen onderwijsinstellingen moord en brand over de toekomst van de schrijfvaardigheid, en misschien zullen AI-docenten uiteindelijk ingrijpend veranderen hoe we leerlingen en studenten lesgeven. Door AI aangedreven amusement maakt het mogelijk dat verhalen voor ons worden gepersonaliseerd, en stuurt schokgolven door Hollywood. En door AI voortgebrachte desinformatie stroomt al door de sociale netwerken op manieren die

moeilijk zijn na te speuren en te bestrijden. Alles staat op het punt ontzettend vreemd te worden. Sterker nog: als je weet waar je moet kijken, wordt alles nu al vreemd.

En dan negeren we nog het grotere probleem, de spreekwoordelijke alien in de kamer. We hebben iets geschapen wat veel slimme mensen het idee bezorgt dat het op een of andere manier de kiem is van een nieuw soort intelligentie. Een AI die nog geen maand nadat ze was uitgevonden slaagde voor zowel de Turingtest ('Kan een computer een mens doen geloven dat hij menselijk is?') als de Lovelace-test ('Kan een computer een mens voor de gek houden bij creatieve werkzaamheden?'), een AI die met vlag en wimpel slaagt voor onze moeilijkste examens, van het behalen van de meestertitel tot het afronden van een opleiding neurochirurgie.⁶ Een AI die topscores behaalt volgens onze hoogste maatstaven voor menselijke creativiteit en tijdens onze beste proeven om bewustzijn te testen. Nog vreemder is dat het niet echt duidelijk is waarom de AI al die dingen kan, ook al hebben we het systeem gebouwd en weten we hoe het technisch gesproken werkt.

19

Niemand weet waar dit allemaal zal toe leiden, ook ik niet. Maar ook al hebben we geen definitieve antwoorden, ik geloof dat ik een nuttige gids kan zijn. Ik heb gemerkt dat ik een invloedrijke stem ben over de implicaties van AI, vooral via mijn nieuwsbrief *One Useful Thing*, ook al ik ben ik zelf geen informaticus. Sterker nog, ik geloof dat het een van mijn voordelen bij het begrijpen van AI is dat ik als docent op de Wharton School langdurig onderzoek heb gedaan naar en heb gepubliceerd over hoe technologieën worden gebruikt. Als gevolg daarvan hebben mijn co-auteurs en ik een paar van de eerste onderzoeken gepubliceerd over AI in het onderwijs en het bedrijfsleven,⁷ terwijl we ook experimenten hebben gedaan met praktische toepassingen van AI die grote AI-bedrijven als voorbeeld hebben genoemd.⁸ Ik spreek regelmatig met organisaties, bedrijven en overheidsinstanties en ook met veel AI-experts om de wereld die we aan het maken zijn te doorgronden. Verder probeer ik de stortvloed aan onderzoeken over dit onderwerp bij te

houden, grotendeels in de vorm van conceptversies van wetenschappelijke artikelen die nog niet het volledige proces van *peer review* hebben doorlopen, maar evengoed waardevolle informatie bieden over dit nieuwe verschijnsel. (Ik zal in dit boek veel citeren uit dit vroege onderzoek om het beeld van de toekomst waarnaar we onderweg zijn aan te vullen, maar daarbij moet je wel bedenken dat het vakgebied zich razendsnel ontwikkelt.) Op grond van al die gesprekken en artikelen kan ik je verzekeren dat niemand volledig kan overzien wat AI behelst, en dat zelfs de mensen die deze systemen maken en gebruiken niet alle implicaties ervan doorgronden.

Daarom zal ik een poging doen om je rond te leiden door AI als een nieuw fenomeen op de wereld, een co-intelligentie, met alle dubbelzinnigheid die dat woord impliceert. We hebben technologieën uitgevonden, van bijlen tot helikopters, die onze fysieke mogelijkheden hebben vergroot, en andere, zoals spreadsheets, waarmee we complexe werkzaamheden hebben geautomatiseerd. Maar we hebben nog nooit een algemeen toepasbare technologie gebouwd die onze intelligentie kan versterken. Nu beschikken mensen over een stuk gereedschap dat kan nabootsen hoe we denken en schrijven, dat zich gedraagt als een co-intelligentie die ons werk kan verbeteren (of vervangen). Maar veel bedrijven die AI ontwikkelen gaan verder, en hopen een bewuste machine te creëren, een werkelijk nieuwe vorm van co-intelligentie die met ons op aarde kan co-existeren. Om greep te krijgen op wat dat betekent, moeten we beginnen bij het begin, met een zeer fundamentele vraag: wat is AI?

Dus daar gaan we beginnen, met een bespreking van de technologie van Large Language Models. Dat ligt ten grondslag aan de vraag hoe wij als mensen het best met deze systemen kunnen samenwerken. Daarna kunnen we bekijken hoe AI ons leven kan veranderen in de rol van collega, docent, expert en zelfs metgezel. Ten slotte kunnen we het hebben over wat dat voor ons zou kunnen betekenen, en wat het betekent om mee te denken met de geest van een alien.

DEEL I

1

DE GEESTEN VAN ALIENS

Praten over AI kan verwarrend zijn, onder meer omdat AI zoveel verschillende dingen heeft betekend, die meestal bij elkaar op één hoop worden gegooid: Siri die jou op bevel een mop vertelt, de Terminator die een schedel verpulvert, algoritmes die kredietwaardigheid voorspellen.

22 We zijn al heel lang gefascineerd door machines die kunnen denken. In 1770 stonden de mensen die hem zagen versteld van de uitvinding van de eerste mechanische schaakcomputer: een schaakbord dat op een ingewikkelde kast stond, terwijl de schaakstukken werden verschoven door een robot die als een Ottomaanse tovenaars was uitgedost. Van 1770 tot 1838 reisde hij de hele wereld rond. Deze machine, die ook wel de Mechanische Turk werd genoemd, versloeg Benjamin Franklin en Napoleon tijdens partijen schaak en zette Edgar Allan Poe, toen hij hem in de jaren 1830 zag, ertoe aan om te speculeren over de mogelijkheid van artificiële intelligentie.¹ Natuurlijk was het allemaal gelogen: tussen de nepraders van de machine zat heel slim een echte schaakmeester verstopt, maar onze bereidheid om te geloven dat machines zouden kunnen denken, hield drie kwart van een eeuw veel grote denkers over de hele wereld voor de gek.

Fast forward naar 1950, toen een stuk speelgoed en een gedachte-experiment, ontwikkeld door verschillende genieën uit het nog altijd piepjonge vak van de informatica, tot een nieuwe opvatting van artificiële intelligentie leidden. Het stuk speelgoed was een in elkaar ge-

knutselde mechanische muis die Theseus heette, ontwikkeld door Claude Shannon, een uitvinder en grappenmaker en de grootste informaticus van de twintigste eeuw.² In een film uit 1950 onthulde hij dat Theseus, aangestuurd door hergebruikte telefoonschakelaars, de weg kon vinden in een ingewikkeld doolhof – het eerste echte voorbeeld van machine learning. Het gedachte-experiment was het imitatiespel, waarbij computerpionier Alan Turing als eerste de theorie uiteenzette dat een machine een functionaliteitsniveau zou kunnen bereiken dat hoog genoeg was om een mens na te bootsen.³ Hoewel computers toen nog een erg nieuwe uitvinding waren, droeg Turgings invloedrijke artikel bij aan het ontstaan van het ontluikende vakgebied van de artificiële intelligentie.

Maar theorieën alleen waren niet genoeg, en een paar van de eerste informatici begonnen programma's te ontwikkelen die de grenzen verlegden van wat al gauw artificiële intelligentie ging heten, een term die in 1956 werd gemunt door John McCarthy, een medewerker van het Massachusetts Institute of Technology (MIT). De vooruitgang ging eerst snel, met computers die werden geprogrammeerd om logische problemen op te lossen en te dammen. Vooraanstaande onderzoekers verwachtten dat een AI binnen tien jaar schaakgrootmeesters zou kunnen verslaan. Maar AI is altijd geplaagd door cycli van hypes, en toen deze beloften niet werden waargemaakt sloeg de ontgoocheling toe, een van de vele 'AI-winters' tijdens welke de vooruitgang van de AI stagneerde en de financiering opdroogde. Er zouden nog meer explosie-en-implosiecycli volgen, waarbij elke explosie gepaard ging met grote technologische vorderingen, zoals artificiële neurale netwerken die de menselijke hersenen nabootsten, gevolgd door ineenstorting wanneer AI de verwachte doelen niet haalde.

De meest recente bloeiperiode van AI begon in de jaren 2010, met de belofte dat de techniek van machine learning kon worden gebruikt voor data-analyse en het doen van voorspellingen. Bij veel van die toepassingen werd een techniek gehanteerd die *supervised learning* wordt genoemd. Dat houdt in dat deze vormen van AI gelabelde