

Hoe AI Werkt

Van Basis tot Begrip — Kunstmatige Intelligentie
Uitgelegd voor Iedereen

2026

Inhoudsopgave

Voorwoord	5
Deel 1: De Basis	7
Hoofdstuk 1: Wat is kunstmatige intelligentie eigenlijk?	8
Hoofdstuk 2: Een korte geschiedenis van AI	14
Hoofdstuk 3: Hoe computers eigenlijk werken	20
Hoofdstuk 4: Data — de brandstof van AI	26
Deel 2: Machine Learning — Leren van Voorbeelden	32
Hoofdstuk 5: Wat is machine learning?	33
Hoofdstuk 6: Patronen herkennen: classificatie en regressie	39
Hoofdstuk 7: Hoe een machine leert: training uitgelegd.....	45
Hoofdstuk 8: Overfitting, underfitting en generalisatie	51
Deel 3: Neurale Netwerken en Deep Learning	57
Hoofdstuk 9: Het brein als inspiratie.....	58
Hoofdstuk 10: Hoe een neuraal netwerk in elkaar zit	65
Hoofdstuk 11: Deep learning: waarom “diep” een verschil maakt	71
Hoofdstuk 12: Backpropagation en gradient descent in mensentaal	77
Deel 4: Taal en AI.....	83
Hoofdstuk 13: Hoe leert een computer taal begrijpen?	84
Hoofdstuk 14: Van woorden naar getallen: embeddings	90
Hoofdstuk 15: Transformers: de doorbraak achter moderne AI	96
Hoofdstuk 16: Attention: waar let een AI op?	102
Hoofdstuk 17: Large Language Models: ChatGPT, Claude, Gemini en co	108
Hoofdstuk 18: Hoe een LLM een antwoord genereert, stap voor stap.....	115
Deel 5: Beeld, Geluid en Meer	122

Hoofdstuk 19: Computer vision: hoe AI leert zien	123
Hoofdstuk 20: Beeldgeneratie: hoe AI plaatjes maakt	129
Hoofdstuk 21: Spraak: luisteren en praten.....	135
Hoofdstuk 22: Aanbevelingssystemen: hoe Netflix en Spotify je “kennen”	141
Deel 6: AI Trainen en Ermee Werken	147
Hoofdstuk 23: Reinforcement learning: leren door vallen en opstaan	148
Hoofdstuk 24: RLHF: hoe AI beleefd en behulpzaam wordt gemaakt	155
Hoofdstuk 25: Prompting: hoe communiceer je goed met AI.....	161
Hoofdstuk 26: Fine-tuning, aangepaste modellen en AI-agents	167
Deel 7: Grenzen, Risico's en Ethiek.....	173
Hoofdstuk 27: Wat AI niet kan (en waarom)	174
Hoofdstuk 28: Hallucinaties: waarom AI dingen verzint.....	180
Hoofdstuk 29: Bias en oneerlijkheid in AI	186
Hoofdstuk 30: Privacy, data en AI.....	192
Hoofdstuk 31: Banen, economie en de impact van AI	198
Hoofdstuk 32: Desinformatie, deepfakes en vertrouwen.....	204
Hoofdstuk 33: Kan AI bewustzijn of gevoelens hebben?.....	210
Deel 8: AI in de Praktijk en de Toekomst	216
Hoofdstuk 34: AI in de gezondheidszorg	217
Hoofdstuk 35: AI in het onderwijs	223
Hoofdstuk 36: AI in het bedrijfsleven en op het werk.....	229
Hoofdstuk 37: AI in wetenschap en onderzoek	235
Hoofdstuk 38: Hoe gebruik je AI verantwoord?	241
Hoofdstuk 39: De toekomst van AI: wat kunnen we verwachten?.....	247
Nawoord	253
Begrippenlijst	254

Voorwoord

Er is een goede kans dat je vandaag al met kunstmatige intelligentie te maken hebt gehad, zonder erbij stil te staan. Misschien ontgrendelde je je telefoon met je gezicht. Misschien stelde je een vraag aan ChatGPT of Claude, kreeg je een routebeschrijving van Google Maps, of liet je Spotify een playlist samenstellen die verrassend goed bij je smaak paste. Kunstmatige intelligentie, meestal afgekort tot AI, is in een paar jaar tijd van iets uit sciencefictionfilms veranderd in een onderdeel van het gewone leven.

En toch blijft het voor de meeste mensen een beetje een zwarte doos. We gebruiken het, we lezen erover in de krant, we horen collega's en familieleden erover praten — maar wat er eigenlijk gebeurt binnen zo'n systeem, hoe het “leert”, en waarom het soms verbluffend slim en soms verrassend dom overkomt, blijft vaak onduidelijk. Dat is niet omdat het te moeilijk is om te begrijpen. Het is omdat de meeste uitleg óf te technisch is (vol wiskunde en jargon, geschreven voor mensen die al programmeren) óf te oppervlakkig (marketingpraat zonder echte inhoud).

Dit boek probeert daar iets tussenin te zitten. Het is geschreven voor mensen zonder enige technische achtergrond — je hoeft nog nooit een regel code te hebben geschreven, en je hoeft geen wiskundeknobbel te hebben. Alles wat in dit boek wordt uitgelegd, wordt uitgelegd met gewone woorden, herkenbare voorbeelden en vergelijkingen uit het dagelijks leven: koken, fietsen leren, een hond trainen, een berg beklimmen met een blinddoek voor. Waar een vakterm onvermijdelijk is, leggen we hem uit op het moment dat hij voor het eerst opduikt, zodat je nooit vastloopt.

Tegelijkertijd doet dit boek geen water bij de wijn wat de inhoud betreft. We slaan de belangrijke concepten niet over omdat ze lastig zijn. In plaats daarvan proberen we ze zo te vertellen dat je ze écht begrijpt — niet alleen de woorden kunt napraten, maar ook kunt uitleggen aan iemand anders waarom een taalmodel soms dingen verzint, of waarom een zelfrijdende auto een voetganger kan herkennen. Aan het einde van dit boek zul je niet in staat zijn om zelf een AI-systeem te bouwen, maar je zult wel precies begrijpen wat er gebeurt als je dat systeem gebruikt — en dat is precies de bedoeling.

Het boek is opgebouwd in acht delen die logisch op elkaar voortbouwen, van de allereerste basis tot de meest actuele toepassingen en vraagstukken. In Deel 1 leggen we de fundamenteën: wat AI eigenlijk is, waar het vandaan komt, en hoe computers en data werken. In Deel 2 en Deel 3 duiken we in machine learning en neurale netwerken — de technieken waarmee computers leren van voorbeelden in plaats van vaste regels. Deel 4 is het hart van het boek: hier leggen we uit hoe taalmodellen zoals ChatGPT, Claude en Gemini werken, van tokens tot transformers tot het moment dat er een antwoord op je scherm verschijnt. Deel 5 laat zien hoe AI ook naar beelden “kijkt” en

met geluid werkt. Deel 6 gaat over hoe deze modellen na hun basistraining verder worden verfijnd, en hoe je er zelf beter mee kunt communiceren. Deel 7 is een eerlijke, genuanceerde blik op de grenzen, risico's en ethische vragen rond AI — want een goed begrip van technologie betekent ook een goed begrip van waar het misgaat. En Deel 8 sluit af met de praktijk: hoe AI wordt gebruikt in de zorg, het onderwijs, het bedrijfsleven en de wetenschap, en wat we — met de nodige bescheidenheid — kunnen verwachten van de toekomst.

Je hoeft dit boek niet per se van kaft tot kaft in één keer te lezen. De delen bouwen op elkaar voort, dus voor het beste begrip is het verstandig om de volgorde aan te houden, zeker de eerste keer. Maar elk hoofdstuk is ook op zichzelf leesbaar, dus als je specifiek wilt weten hoe beeldgeneratie werkt of waarom AI soms dingen verzint, kun je gerust direct naar dat hoofdstuk springen. Achterin vind je een begrippenlijst waarin alle belangrijke termen nog eens kort op een rijtje staan, handig om op terug te grijpen.

Eén ding vooraf: AI is geen magie, en het is ook geen bewust wezen dat denkt zoals wij. Het is, hoe indrukwekkend de resultaten soms ook zijn, in de kern wiskunde en statistiek — patronen die uit enorme hoeveelheden data zijn gehaald door mensen die deze systemen hebben gebouwd. Zodra je dat eenmaal doorziet, wordt AI een stuk minder mysterieus, en een stuk makkelijker om verstandig mee om te gaan. Dat is precies wat dit boek je wil meegeven: niet angst, niet blind enthousiasme, maar begrip.

Veel leesplezier.

Deel 1: De Basis

Hoofdstuk 1: Wat is kunstmatige intelligentie eigenlijk?

Stel je voor: je pakt 's ochtends je telefoon, en zonder dat je erbij nadenkt gebeurt er van alles wat twintig jaar geleden nog science fiction leek. Je telefoon herkent je gezicht en ontgrendelt zichzelf. Je e-mailprogramma heeft die ene mail met “GRATIS!!! Klik hier voor je miljoenenwinst” al keurig in de spammap gestopt, zonder dat jij er ook maar naar hebt gekeken. Onderweg naar je werk vertelt je navigatie-app precies welke route het snelst is, omdat het duizenden andere auto's op de weg in de gaten houdt. En 's avonds, als je op de bank ploft, stelt Netflix je een serie voor die verdacht goed bij je smaak past. Al deze dingen — gezichtsherkenning, spamfilters, verkeersvoorspelling, aanbevelingssystemen — zijn voorbeelden van kunstmatige intelligentie, oftewel AI. En de kans is groot dat je er, zonder het je te realiseren, elke dag tientallen keren mee in aanraking komt.

Maar wat is die kunstmatige intelligentie dan eigenlijk, in de kern? Laten we beginnen met wat het niet is. Als je het woord “AI” hoort, denk je misschien meteen aan robots uit films: HAL 9000 die weigert een luik te openen, Terminator die de wereld probeert over te nemen, of een androïde die zich bewust wordt van zijn eigen bestaan en zich tegen zijn makers keert. Dat is een prachtig beeld voor een film, maar het heeft weinig te maken met de AI die vandaag de dag daadwerkelijk bestaat. Er is geen sprake van een computer die “wakker wordt”, gevoelens ontwikkelt, of stiekem plannen smeedt. Wat we vandaag kunstmatige intelligentie noemen, is iets veel minder dramatisch en veel praktischer: het is software — computerprogramma's — die taken uitvoert waarvoor normaal gesproken menselijke intelligentie nodig zou zijn. Denk aan dingen als een gezicht herkennen, een gesproken zin verstaan, een tekst vertalen, een schaakpartij winnen, of een vraag beantwoorden. Dat zijn allemaal taken die vroeger uitsluitend door mensen (of dieren) gedaan konden worden, en die nu, dankzij slimme programma's, ook door computers gedaan kunnen worden.

Het cruciale woord hier is “taken”. AI is geen alwetend brein dat overal verstand van heeft. Het is, in de meeste gevallen, een gereedschap dat is toegesneden op één specifiek klusje. Om dat verschil helder te maken, gebruiken onderzoekers twee begrippen: smalle AI en algemene AI.

Smalle AI — in het Engels “narrow AI” — is verreweg de meest voorkomende vorm, en het is de enige vorm die op dit moment werkelijk bestaat. Smalle AI is fantastisch goed in precies één ding, en helemaal nergens anders in. Denk aan een schaakcomputer: die kan de beste schaakspelers ter wereld inmaken, moeiteloos, elke keer weer. Maar vraag diezelfde schaakcomputer om een boodschappenlijstje te maken, of om uit te leggen waarom je oma boos werd op het verjaardagsfeestje, en hij staat compleet met de mond vol tanden — hij “snapt” alleen schaken, en dan ook nog alleen de spelregels en de zetten, niet het idee van een spelletje, van winnen, van plezier. Je spamfilter is precies zo: het is uitstekend in het herkennen van

rommelmail, maar het kan niet autorijden, geen recept bedenken, en geen gesprek voeren. Gezichtsherkenning op je telefoon kan gezichten van elkaar onderscheiden met verbluffende precisie, maar diezelfde technologie kan niet ineens ook stemmen herkennen of een auto besturen — daar is weer een heel andere, apart gebouwde en apart getrainde AI voor nodig. Elk van deze systemen is dus eigenlijk een specialist met tunnelvisie: adembenemend goed binnen zijn eigen kleine wereldje, en volkomen hulpeloos daarbuiten.

Ook de chatbots die de laatste jaren zo enorm populair zijn geworden — zoals ChatGPT of Claude, de AI-assistent van het bedrijf Anthropic — vallen onder smalle AI, ook al voelt dat misschien verrassend. Zulke chatbots kunnen namelijk over een verbluffend breed scala aan onderwerpen praten: ze kunnen een gedicht schrijven, een programmeerfout opsporen, een historische gebeurtenis uitleggen, en advies geven over een moeilijk gesprek met je baas. Dat lijkt dus helemaal niet zo smal. Maar het gaat hier nog steeds, ten diepste, om één taak: het voorspellen en genereren van tekst (en soms afbeeldingen of geluid) op basis van patronen die het systeem tijdens zijn training heeft gezien — daarover leer je in de volgende delen van dit boek veel meer. Zo'n chatbot heeft geen eigen wil, geen bewustzijn, geen honger of verveling, en geen begrip van de wereld zoals jij dat hebt. Het is, hoe indrukwekkend de output ook oogt, nog steeds een zeer geavanceerd stappenplan dat tekst produceert — geen denkend wezen.

Dat is meteen ook de reden waarom je in dit boek nooit zult lezen dat een chatbot “denkt” of “wil” zoals een mens dat doet, ook al gebruiken zelfs mensen die in de AI-industrie werken soms losjes zulke menselijke woorden, simpelweg omdat we nu eenmaal geen betere woorden hebben om te beschrijven wat het systeem doet. Wanneer je in de rest van dit boek woorden tegenkomt als “leren”, “begrijpen” of “beslissen” toegepast op een computer, bedoelen we daarmee altijd het technische, sterk vereenvoudigde equivalent — een wiskundig proces dat op sommige punten lijkt op wat er in een mensenbrein gebeurt, maar dat op geen enkele bewezen manier gepaard gaat met een innerlijke ervaring, een gevoel, of een bewustzijn zoals jij dat kent. Dat onderscheid blijf je in dit boek voortdurend tegenkomen, en het is de moeite waard om het nu al scherp te hebben, want het voorkomt een hoop misverstanden verderop.

Daartegenover staat het idee van algemene AI, ook wel “AGI” genoemd (van het Engelse “Artificial General Intelligence”). Dit zou een kunstmatige intelligentie zijn die, net als een mens, in principe alles zou kunnen leren en toepassen: de ene dag een schaakpartij winnen, de volgende dag een auto repareren, daarna een grap vertellen, een nieuwe wiskundige stelling bedenken, troost bieden aan een verdrietige vriend, en 's avonds nog even een recept improviseren met de restjes uit de koelkast — allemaal met hetzelfde ene “brein”. Zo'n systeem zou niet, zoals de huidige AI, speciaal getraind hoeven worden voor elke afzonderlijke taak, maar zou zichzelf flexibel

kunnen aanpassen aan nieuwe, onbekende situaties, ongeveer zoals een mens dat doet. Het belangrijkste om te onthouden is dit: algemene AI bestaat op dit moment niet. Het is een doel waar sommige onderzoekers naar streven, en waar in de media en door bedrijven soms enthousiast (of juist angstig) over gespeculeerd wordt, maar niemand heeft tot nu toe een systeem gebouwd dat werkelijk zo veelzijdig is als een mens. Alles wat je vandaag aan AI tegenkomt — hoe indrukwekkend ook — is nog steeds smalle AI, hooguit een smalle AI die toevallig heel breed inzetbaar is binnen het domein van taal en tekst.

Waarom is dit onderscheid zo belangrijk om te begrijpen, al helemaal aan het begin van dit boek? Omdat het je helpt om AI met heldere ogen te bekijken, in plaats van door de bril van films en sciencefictionromans. Wanneer je begrijpt dat AI-systemen gespecialiseerde gereedschappen zijn — hoe knap ook gemaakt — dan besef je ook meteen waarom ze soms zulke rare, bijna kinderlijke fouten maken. Een chatbot die een feit verzint dat helemaal niet klopt, doet dat niet omdat hij liegt of iets verbergt, maar omdat hij, simpel gezegd, geen besef heeft van “waar is” op de manier waarop jij dat hebt. Hij is een taalvoorspeller, geen encyclopedie met een geweten. Dat besef helpt je ook om AI beter te gebruiken: je weet dan wanneer je een systeem kunt vertrouwen (bijvoorbeeld bij het herkennen van een gezicht, waar de technologie extreem betrouwbaar in is geworden) en wanneer je extra voorzichtig moet zijn (bijvoorbeeld bij het controleren van feiten die een chatbot je vertelt).

Voordat we verder gaan, is het de moeite waard om nog even stil te staan bij dat woord “intelligentie” zelf, want het is eigenlijk een lastig woord. Wat betekent het precies om intelligent te zijn? Is het slim rekenen, snel schaken, mooi schrijven, goed inschatten hoe een ander zich voelt, of juist creatief een probleem oplossen waar niemand eerder aan gedacht heeft? Mensen zijn het onderling al niet eens over wat menselijke intelligentie precies inhoudt, en dat maakt het des te lastiger om te bepalen wanneer een machine “intelligent” genoemd mag worden. Onderzoekers hebben daarom, in plaats van een sluitende definitie te zoeken, vooral gekeken naar gedrag: als een systeem een taak uitvoert die, wanneer een mens het zou doen, we intelligent zouden noemen, dan noemen we dat systeem in de volksmond en in dit boek gemakshalve ook “kunstmatig intelligent” — ook al is er geen enkele garantie dat er van binnen iets gebeurt dat lijkt op menselijk denken. Dat is een belangrijke nuance om vast te houden: AI wordt beoordeeld op wat het doet, niet op wat het “is” of “ervaart”. Een schaakcomputer die een grootmeester verslaat, wordt intelligent genoemd om zijn prestatie, niet omdat er iemand “binnenin” zit die plezier beleeft aan het winnen.

Laten we nog wat meer voorbeelden uit het dagelijks leven bij elkaar zetten, want die maken het concept veel tastbaarder dan een definitie ooit kan. Je spamfilter — genoemd aan het begin van dit hoofdstuk — bekijkt binnenkomende e-mails en beslist op basis van patronen (bepaalde woorden, de afzender, de opmaak) of iets

waarschijnlijk rommel is. Gezichtsherkenning op je telefoon vergelijkt de vorm van je gezicht met een eerder opgeslagen “sjabloon” en beslist binnen een fractie van een seconde of jij het echt bent. Netflix en soortgelijke diensten houden bij wat jij (en miljoenen anderen) hebben bekeken, en gebruiken die patronen om te voorspellen wat je waarschijnlijk ook leuk zou vinden — een soort digitale vriend die zegt: “mensen die hield van X, hielden ook vaak van Y.” Navigatie-apps zoals Google Maps of Waze verzamelen voortdurend locatiegegevens van miljoenen telefoons tegelijk, zien waar het verkeer langzamer rijdt dan normaal, en berekenen op basis daarvan de snelste route voor jou. Zelfs de manier waarop je bank verdachte transacties opspoort — bijvoorbeeld wanneer er ineens een aankoop wordt gedaan in een land waar jij nog nooit bent geweest — draait vaak op AI die patronen in jouw uitgavengedrag heeft leren herkennen. En natuurlijk de chatbots: systemen als ChatGPT en Claude die met je kunnen “praten” alsof je met een heel goed geïnformeerde gesprekspartner zit te chatten, terwijl ze in werkelijkheid, woord voor woord, voorspellen wat een zinnig antwoord zou kunnen zijn.

Er zijn nog meer voorbeelden die het waard zijn om te noemen, omdat ze laten zien hoe breed AI inmiddels is doorgedrongen in gewone, alledaagse technologie. Wanneer je op je telefoon typt en er alvast een suggestie voor het volgende woord verschijnt, of wanneer een tikfout automatisch wordt gecorrigeerd, is dat AI aan het werk. Wanneer je tegen een slimme speaker in je huiskamer zegt “zet het licht uit” en de lamp daadwerkelijk dooft, dan heeft er eerst spraakherkenning plaatsgevonden — jouw gesproken woorden zijn omgezet in tekst en vervolgens geïnterpreteerd als een opdracht — allemaal dankzij AI. Wanneer een ziekenhuis een röntgenfoto laat meekijken door software die verdachte vlekjes markeert die een radioloog zou kunnen missen, is dat een vorm van AI die medische professionals ondersteunt. Wanneer een vertaal-app een menukaart in een vreemde taal in real time voor je vertaalt terwijl je hem met je camera bekijkt, zie je beeldherkenning en taalverwerking naadloos samenwerken. En wanneer sociale media bepalen welke berichten bovenaan jouw tijdlijn verschijnen, is ook dat het resultaat van een AI-systeem dat, op basis van jouw eerdere gedrag, voorspelt wat jouw aandacht het langst zal vasthouden. Het is die laatste categorie — aanbevelingssystemen die precies weten wat onze aandacht trekt — die overigens ook laat zien dat AI niet per definitie alleen maar handig of onschuldig is; we komen in een later deel van dit boek uitgebreid terug op de vraag welke risico’s en dilemma’s deze technologie met zich meebrengt.

Al deze voorbeelden hebben iets gemeen: het zijn systemen die, in plaats van precies te doen wat een mens ze letterlijk heeft voorgeschreven, patronen hebben geleerd uit heel veel voorbeelden. Dat onderscheid — tussen een computer die exact volgt wat een programmeur heeft opgeschreven, en een computer die zelf patronen ontdekt in data — is misschien wel het allerbelangrijkste idee in dit hele boek, en we gaan er in hoofdstuk 3 uitgebreid op in. Voor nu is het genoeg om te onthouden: AI is geen magie, geen bewustzijn, en geen sciencefictionrobot. Het is knap gemaakte software

die, op basis van patronen die het geleerd heeft uit grote hoeveelheden voorbeelden, taken uitvoert die vroeger alleen mensen konden.

En daarmee kom je meteen bij de vraag waarom dit boek de moeite waard is om te lezen. AI is niet langer iets voor specialisten in witte jassen achter dikke computerschermen. Het zit in je telefoon, in je auto, in je bank, in je werk, in het nieuws, en steeds vaker ook in de beslissingen die overheden en bedrijven over jou nemen. Je hoeft geen programmeur te worden om te begrijpen hoe het werkt — net zoals je geen automonteur hoeft te zijn om te snappen dat een auto rijdt dankzij een motor die brandstof verbrandt. Maar een beetje basiskennis maakt een enorm verschil. Het helpt je onderscheid te maken tussen realistische verwachtingen en overdreven beloftes (of overdreven angst). Het helpt je begrijpen waarom een chatbot soms fouten maakt, en waarom die fouten helemaal geen “bugs” zijn maar een logisch gevolg van hoe het systeem in elkaar zit. Het helpt je mee te praten over een technologie die, of je het nu wilt of niet, een steeds grotere rol gaat spelen in hoe de wereld om je heen functioneert. In de rest van dit boek bouwen we, stap voor stap en zonder ingewikkeld jargon, precies dat begrip op: van de allereerste ideeën over denkende machines, via de manier waarop computers en data werken, tot de moderne technieken achter de AI-systemen die vandaag ons leven binnendringen. Je hoeft nergens voorkennis voor te hebben. Je hoeft alleen nieuwsgierig te zijn — en dat ben je, anders had je dit boek niet opengeslagen.

Wat je nu weet

- AI van vandaag is 'smalle AI': knap gereedschap voor specifieke taken, geen bewust digitaal brein zoals in films.
- Chatbots zoals ChatGPT en Claude vallen ook onder smalle AI: ze doen in de kern één ding — tekst voorspellen en genereren op basis van geleerde patronen.
- Het onderscheid tussen 'vaste regels programmeren' en 'patronen leren uit voorbeelden' is de rode draad van dit hele boek.

Hoofdstuk 2: Een korte geschiedenis van AI

Om te begrijpen waar de AI van vandaag vandaan komt, moeten we terug in de tijd — naar een periode ver voordat er ook maar één computer op een bureau stond, laat staan in een broekzak. Het verhaal begint bij een Britse wiskundige met een buitengewoon scherpe geest: Alan Turing.

In 1950, midden in de nasleep van de Tweede Wereldoorlog waarin Turing zelf een cruciale rol had gespeeld bij het kraken van Duitse geheime codes, schreef hij een artikel met een simpele maar diepgaande vraag: kunnen machines denken? Turing beseftte al snel dat die vraag lastig te beantwoorden is, want wat “denken” precies betekent, is filosofisch gezien een moeras. Dus bedacht hij iets slims: in plaats van te proberen te definiëren wat denken is, stelde hij een praktische test voor. Laat een mens via een tekstscherf — zonder de gesprekspartner te zien of horen — gesprekken voeren, soms met een computer en soms met een ander mens. Als die mens niet kan raden wie van de twee de computer is, dan, zo redeneerde Turing, zouden we de machine op zijn minst moeten toegeven dat hij zich intelligent gedraagt. Deze gedachte-experiment staat sindsdien bekend als de Turingtest, en hoewel wetenschappers het er tegenwoordig over eens zijn dat de test niet perfect is — een systeem kan mensen om de tuin leiden zonder werkelijk te “begrijpen” wat het zegt, zoals we in hoofdstuk 1 al zagen — was het idee baanbrekend. Turing plantte het zaadje voor een compleet nieuw onderzoeksveld, jaren voordat er zelfs maar een woord voor bestond.

Turing zelf zou de vruchten van zijn eigen inzichten helaas nooit zien rijpen: hij overleed al in 1954, nog geen vier jaar na het schrijven van zijn beroemde artikel, en zijn leven eindigde tragisch — hij werd in eigen land vervolgd vanwege zijn homoseksualiteit, destijds strafbaar in het Verenigd Koninkrijk, ondanks (of misschien wel deels dankzij het stilzwijgen rond) zijn cruciale bijdrage aan de geallieerde overwinning in de oorlog. Pas decennia later kreeg hij de erkenning die hij verdiende, en tegenwoordig wordt hij algemeen gezien als een van de grondleggers van zowel de informatica als de kunstmatige intelligentie.

In de jaren direct na Turings artikel begonnen ook de eerste, nog heel eenvoudige experimenten met kunstmatig “denkende” machines. Een van de bekendste vroege voorbeelden was een programma genaamd ELIZA, gebouwd in de jaren zestig door de informaticus Joseph Weizenbaum. ELIZA deed alsof hij een psychotherapeut was: hij herkende een paar sleutelwoorden in wat een gebruiker typte en gooide die woorden er, in de vorm van een wedervraag, weer uit — typte je “ik voel me verdrietig over mijn moeder”, dan antwoordde ELIZA misschien iets als “vertel me eens meer over je moeder.” Het programma had geen enkel begrip van verdriet, moeders, of wat dan ook; het volgde simpele, van tevoren geschreven trucjes om tekst te herschikken. Toch waren sommige testgebruikers, tot Weizenbaums eigen verbazing en ongenoegen,

oprecht van slag toen ze hoorden dat ze niet met een echt begripvol wezen hadden gesproken — een vroeg signaal van hoe gemakkelijk mensen geneigd zijn om betekenis en begrip toe te dichten aan een machine die enkel woorden slim herschikt. Dat fenomeen duikt, zoals je verderop in dit boek zult lezen, tot op de dag van vandaag op bij chatbotgebruikers.

Dat woord kwam er zes jaar later. In de zomer van 1956 kwam een klein groepje wetenschappers samen op de universiteit van Dartmouth, in de Amerikaanse staat New Hampshire, voor wat achteraf een van de invloedrijkste bijeenkomsten uit de geschiedenis van de wetenschap zou blijken. De organisator, een jonge onderzoeker genaamd John McCarthy, gebruikte in de uitnodiging voor het eerst officieel de term “artificial intelligence” — kunstmatige intelligentie. De groep, met onder anderen ook latere grootheden als Marvin Minsky, geloofde oprecht dat als je maar de juiste slimme mensen twee maanden lang bij elkaar zette, ze substantiële vooruitgang zouden kunnen boeken richting machines die net zo konden redeneren, taal gebruiken en problemen oplossen als mensen. Dat optimisme was, laten we het vriendelijk zeggen, behoorlijk overmoedig — het zou uiteindelijk niet twee maanden duren, maar tientallen jaren, en het doel is tot op de dag van vandaag nog niet bereikt. Maar de bijeenkomst gaf het vakgebied wel zijn naam, zijn eerste onderzoekers, en een golf van enthousiasme.

Die eerste decennia van AI-onderzoek, de jaren zestig en een deel van de jaren zeventig, stonden bol van hooggespannen verwachtingen. Onderzoekers voorspelden serieus dat machines binnen een paar decennia menselijk niveau van intelligentie zouden evenaren. Er kwamen programma's die eenvoudige wiskundige problemen konden oplossen, die konden “praten” in beperkte gesprekken, en die simpele spelletjes konden spelen. Voor die tijd was dat verbluffend. Maar de techniek liep al snel tegen zijn grenzen aan: computers waren traag en hadden nauwelijks geheugen, en de problemen die onderzoekers écht wilden oplossen — een zin in gewone taal begrijpen, een foto herkennen, gezond verstand toepassen — bleken vele malen lastiger dan gedacht. De vooruitgang stokte, de beloftes werden niet waargemaakt, en overheden en investeerders die miljoenen hadden gestoken in AI-onderzoek begonnen ongeduldig te worden. Rond het midden van de jaren zeventig droogde de financiering grotendeels op. Deze periode van teleurstelling en bezuinigingen wordt door historici een “AI-winter” genoemd — een treffende metafoor voor een tijd waarin het onderzoeksveld, na een hoopvolle lente vol beloftes, in de kou kwam te staan.

Maar zoals dat vaker gaat met winters, kwam er ook weer een lente. In de jaren tachtig beleefde AI een heropleving, ditmaal gedreven door zogeheten expertsystemen. Het idee achter een expertsysteem was relatief eenvoudig: haal de kennis van een menselijke expert — bijvoorbeeld een arts die longziekten diagnosticeert, of een geoloog die mineralen herkent — en giet die kennis in een lange reeks als-dan-regels die een computer kan volgen. “Als de patiënt koorts heeft én

hoest al langer dan twee weken duurt én er is bloed bij het slijm, overweeg dan diagnose X.” Een van de bekendste voorbeelden was een systeem genaamd MYCIN, ontwikkeld aan de universiteit van Stanford, dat artsen hielp bij het voorschrijven van de juiste antibiotica op basis van honderden zorgvuldig met de hand ingevoerde medische regels — en dat in tests soms zelfs betere adviezen gaf dan ervaren artsen. Bedrijven zagen hier commercieel goud in en investeerden opnieuw fors. Voor een tijdje leek het alsof AI eindelijk zijn beloftes zou waarmaken, met systemen die daadwerkelijk in fabrieken en ziekenhuizen werden gebruikt. Maar ook expertsystemen bleken uiteindelijk te star: ze werkten alleen goed binnen het smalle domein waarvoor ze precies waren gebouwd, ze waren peperduur om te onderhouden — elke nieuwe regel moest met de hand worden ingevoerd door een specialist — en ze konden niet omgaan met situaties die de bouwers niet hadden voorzien. Tegen het einde van de jaren tachtig kwam de tweede grote AI-winter, opnieuw met inkrimpende budgetten en een golf van scepsis over of “kunstmatige intelligentie” ooit meer zou worden dan een te grote belofte.

Wat de derde en blijvende doorbraak uiteindelijk mogelijk maakte, was een fundamentele verschuiving in hoe onderzoekers over het probleem dachten. In plaats van te proberen alle kennis en regels handmatig in een computer te stoppen — zoals bij expertsystemen — begonnen wetenschappers vanaf de jaren negentig en vooral de jaren tweeduizend steeds meer te experimenteren met machine learning: technieken waarbij een computer zelf patronen leert herkennen door naar heel veel voorbeelden te kijken, in plaats van dat een mens alle regels met de hand uitschrijft (dit verschil werken we in het volgende hoofdstuk helemaal uit). Deze aanpak bestond in de kern eigenlijk al sinds de jaren vijftig en zestig — al vroeg experimenteerden onderzoekers met zogeheten neurale netwerken, wiskundige systemen die losjes geïnspireerd zijn op hoe hersencellen met elkaar communiceren — maar ze werkten in de praktijk matig, simpelweg omdat de computers van die tijd te traag waren en er nog veel te weinig digitale data beschikbaar was om iets zinnigs van te leren. Twee ontwikkelingen veranderden dat geleidelijk: computers werden exponentieel sneller en krachtiger, en de opkomst van het internet zorgde voor een lawine aan digitale data — foto’s, teksten, klikgedrag — waar systemen van konden leren.

Het moment waarop dit alles echt samenkwam met een knal, was 2012. Dat jaar werd er een jaarlijkse wedstrijd gehouden, ImageNet genaamd, waarbij computersystemen van over de hele wereld streden om foto’s zo goed mogelijk automatisch te herkennen en te categoriseren — is dit een kat, een fiets, een berg? Jarenlang boekten de deelnemende systemen langzame, geleidelijke vooruitgang. Maar in 2012 kwam een team met een systeem dat gebruikmaakte van een specifieke vorm van neurale netwerken, “deep learning” genoemd — “diep leren”, verwijzend naar netwerken met veel opeenvolgende lagen van kunstmatige rekeneenheden. Dat systeem versloeg de concurrentie niet met een klein verschil, maar met een verpletterende marge: het aantal fouten daalde in één klap drastisch ten opzichte van het beste systeem van het

jaar ervoor. De hele onderzoeksgemeenschap was met stomheid geslagen. Binnen een paar jaar schakelde vrijwel iedereen die met beeldherkenning bezig was over op deep learning, en al snel breidde die aanpak zich uit naar spraakherkenning, vertaalsystemen, en tal van andere toepassingen. Dit was het moment waarop AI, na decennia van pieken en dalen, definitief zijn opmars begon die tot op vandaag doorzet.

Twee dingen maakten deze doorbraak in 2012 precies op dát moment mogelijk, en het is leerzaam om ze te noemen, omdat je ze in de rest van dit boek nog vaak zult tegenkomen. Ten eerste waren computers, en met name een bepaald type rekenchip dat oorspronkelijk was ontworpen om videogames er soepel uit te laten zien (een grafische kaart, ook wel GPU genoemd), inmiddels krachtig genoeg geworden om de miljarden benodigde berekeningen voor een diep neurale netwerk in een redelijke tijd uit te voeren — iets wat een paar jaar eerder simpelweg onbetaalbaar traag zou zijn geweest. Ten tweede was er, dankzij het inmiddels volwassen geworden internet, eindelijk genoeg data beschikbaar: de wedstrijd waarin het winnende systeem uitblonk, gebruikte een verzameling van meer dan een miljoen zorgvuldig gelabelde foto's om van te leren. Zonder die twee ontwikkelingen — genoeg rekenkracht, genoeg data — was diezelfde technische aanpak, die in de kern al sinds de jaren zestig bekend was, nooit tot bloei gekomen. Het is een patroon dat je door de hele geschiedenis van AI heen ziet terugkeren: goede ideeën lagen vaak al jaren op de plank, maar wachtten op het moment waarop de technologie eromheen — rekenkracht, geheugen, data — eindelijk volwassen genoeg was om ze werkelijk te laten werken.

Er was nog één cruciale uitvinding nodig om te komen bij de chatbots die je vandaag de dag kent. In 2017 publiceerde een groep onderzoekers van Google een wetenschappelijk artikel met de speelse titel “Attention Is All You Need” — “aandacht is alles wat je nodig hebt”. Daarin beschreven ze een nieuwe architectuur, een nieuwe manier om een neurale netwerk op te bouwen, die ze de “transformer” noemden. Zonder op de technische details vooruit te lopen (die komen uitgebreid aan bod in latere delen van dit boek): de transformer bleek buitengewoon goed in het verwerken van taal, omdat hij, kort gezegd, tegelijkertijd naar een hele zin (of zelfs een heel document) kon “kijken” en kon inschatten welke woorden voor elkaar belangrijk zijn, in plaats van woord voor woord in een strikte volgorde te werken zoals eerdere technieken deden. Deze architectuur bleek zich bovendien uitstekend te lenen voor enorme schaalvergroting: hoe meer data en rekenkracht je erin stopte, hoe beter de resultaten leken te worden, zonder dat er een duidelijk plafond in zicht kwam. De transformer werd de motor achter vrijwel alle grote taalmodellen die daarna volgden, inclusief de systemen achter ChatGPT en Claude.

Tussen die uitvinding in 2017 en de wereldwijde doorbraak eind 2022 zaten nog een paar jaar van gestage, maar voor de buitenwereld grotendeels onopgemerkte vooruitgang. Het bedrijf OpenAI bracht in die tussenliggende jaren een reeks steeds

grotere taalmodellen uit, elk gebouwd op de transformer-architectuur en elk getraind op een steeds grotere berg tekst met steeds meer rekenkracht. De eerste versies waren voor het grote publiek nauwelijks bekend en werden vooral door onderzoekers en ontwikkelaars getest; ze konden al aardig lopende tekst produceren, maar maakten ook nog geregeld vreemde, evident onlogische fouten. Met elke nieuwe, grotere versie werden de resultaten echter merkbaar beter, tot het punt waarop de kwaliteit van de gegenereerde tekst een drempel overschreed die het geschikt maakte voor een breed publiek, niet langer alleen voor technici die de eigenaardigheden en beperkingen van het systeem al kenden.

De laatste stap in dit verhaal — en degene die AI van een onderwerp voor technici veranderde in wereldnieuws — vond plaats eind 2022. Het bedrijf OpenAI bracht toen een chatbot uit, ChatGPT, gebouwd op deze transformer-technologie en getraind op enorme hoeveelheden tekst van het internet. Wat volgde was ongekend: binnen enkele dagen gebruikten miljoenen mensen wereldwijd het systeem, en binnen twee maanden waren dat er meer dan honderd miljoen — de snelst groeiende consumententoepassing in de geschiedenis tot op dat moment. Voor het eerst konden gewone mensen, zonder enige technische kennis, rechtstreeks een gesprek voeren met een AI-systeem dat vloeiend kon schrijven, uitleggen, samenvatten en brainstormen, in bijna elke taal en over bijna elk onderwerp. Dit “ChatGPT-moment” zorgde voor een schokgolf: bedrijven over de hele wereld begonnen AI-strategieën te schrijven, scholen worstelden ineens met de vraag hoe om te gaan met leerlingen die hun huiswerk lieten schrijven door een chatbot, journalisten schreven de ene na de andere voorpagina's, en concurrenten — waaronder Google en het bedrijf Anthropic, dat de chatbot Claude ontwikkelde — brachten in hoog tempo hun eigen versies uit. Wat decennia lang het domein was geweest van onderzoekers, sciencefictionschrijvers en af en toe een optimistische krantenkop, werd in een paar maanden tijd iets waar je oma, je buurman en je collega's dagelijks mee omgingen.

Als je dit hele verhaal overziet — van Turings voorzichtige vraag in 1950, via de euforie en teleurstelling van de eerste decennia, de opkomst en ondergang van de expertsystemen, tot de stille revolutie van machine learning en de explosieve doorbraak van deep learning en de transformer — dan zie je een patroon dat de geschiedenis van AI kenmerkt: golven van euforisch optimisme, gevolgd door periodes van teleurstelling, telkens weer gevolgd door een onverwachte doorbraak die alles verandert. Het is goed om dat patroon in je achterhoofd te houden terwijl je de rest van dit boek leest — en terwijl je de komende jaren het nieuws over AI blijft volgen. De geschiedenis leert ons zowel voorzichtigheid als verwondering: voorzichtigheid, omdat beloftes over AI in het verleden regelmatig te groot bleken; en verwondering, omdat de technologie, ondanks alle tegenslagen, uiteindelijk toch keer op keer verder is gekomen dan wie dan ook voor mogelijk hield.

Wat je nu weet

- Alan Turing stelde in 1950 de vraag 'kunnen machines denken?' en bedacht de Turingtest als praktische toets.
- De geschiedenis van AI verloopt in golven: euforie, teleurstelling ('AI-winters'), en onverwachte doorbraken die alles weer veranderen.
- Dat patroon leert zowel voorzichtigheid (beloftes vielen vaker tegen dan gedacht) als verwondering (de vooruitgang ging uiteindelijk verder dan verwacht).

Hoofdstuk 3: Hoe computers eigenlijk werken

Voordat we dieper kunnen duiken in hoe AI precies leert en denkt — of beter gezegd, hoe het doet alsof het denkt — moeten we een stap terugzetten en kijken naar iets nog fundamentele: wat een computer eigenlijk is, en hoe hij in de basis werkt. Dit hoofdstuk is bewust helemaal bij het begin begonnen, zonder enige aanname van voorkennis, omdat dit de brug vormt naar alles wat daarna komt. Zodra je dit begrijpt, valt de rest van het boek op zijn plaats.

Laten we beginnen met de meest simpele, maar ook meest fundamentele omschrijving die er is: een computer is een machine die instructies razendsnel en met absolute precisie uitvoert. Dat is het, in essentie. Geen magie, geen intelligentie in de menselijke zin, geen begrip. Een computer weet niet wát hij doet in de manier waarop jij dat weet — hij voert simpelweg, stap voor stap, een lijst instructies uit, en hij doet dat zo verbluffend snel (we hebben het over miljarden stappen per seconde bij een gewone laptop) dat het resultaat op ons, mensen, indrukwekkend en soms zelfs “slim” overkomt. Vergelijk het met een extreem gedisciplineerde keukenhulp die geen enkel eigen initiatief neemt, maar een recept tot in de puntjes en zonder ook maar één keer af te wijken kan uitvoeren, en dat in een oogwenk, miljoenen keren achter elkaar. Die snelheid en precisie, gecombineerd, zijn de bron van bijna alles wat een computer voor ons kan betekenen.

Maar hoe “praat” je eigenlijk tegen zo’n machine? Hier komen we bij een van de meest fundamentele bouwstenen van de hele digitale wereld: de bit. Het woord “bit” is een samentrekking van “binary digit”, ofwel “binair cijfer” — een cijfer dat maar twee waarden kan aannemen: nul of één. Waarom werkt alles in een computer met slechts die twee simpele waarden, in plaats van bijvoorbeeld met de tien cijfers die wij mensen gewend zijn (0 tot en met 9)? Het antwoord heeft te maken met hoe een computer fysiek in elkaar zit. Diep in de kern bestaat een computer uit miljarden minuscule elektronische schakelaartjes. Een schakelaar kan, net als een lichtschakelaar aan de muur van je huiskamer, maar twee standen hebben: aan of uit. Aan staat gelijk aan een één, uit staat gelijk aan een nul. Door miljarden van die schakelaartjes op ingenieuze manieren met elkaar te combineren, kan een computer met louter nullen en enen toch ontzettend complexe dingen weergeven en berekenen: getallen, letters, kleuren, geluid, video — werkelijk alles wat je op een scherm ziet of uit een speaker hoort, is uiteindelijk teruggebracht tot lange, lange reeksen van nullen en enen. Een simpele foto op je telefoon bijvoorbeeld bestaat uit miljoenen piepkleine puntjes, pixels genoemd, en de kleur van elke pixel wordt vastgelegd als een setje bits. Zet die miljoenen setjes bij elkaar en je krijgt, na verwerking, een foto van je hond op het scherm. Het is een beetje zoals morsecode, waarbij elke letter wordt weergegeven met combinaties van korte en lange signalen — alleen dan met nog maar twee

mogelijke signalen (aan/uit) in plaats van kort/lang, en op een schaal die miljoenen keren groter en sneller is.

Hetzelfde geldt voor tekst. Elke letter die je nu leest, wordt in het geheugen van een computer opgeslagen als een vaste combinatie van nullen en enen — ergens is afgesproken dat bijvoorbeeld de hoofdletter “A” wordt weergegeven door een specifiek setje van acht bits, de letter “B” door een ander specifiek setje, enzovoort, voor elke letter, cijfer en leesteken die je kent (en ook voor duizenden tekens uit andere alfabetten en zelfs emoji’s). Zo’n setje van acht bits noemen we, niet toevallig, een byte, en de meeste bestandsgroottes die je weleens voorbij ziet komen — kilobytes, megabytes, gigabytes — zijn simpelweg tellingen van hoeveel van die setjes bij elkaar nodig zijn om iets op te slaan: hoe groter het bestand, hoe meer nullen en enen erin verstopt zitten. Muziek werkt volgens hetzelfde principe: een geluidsgolf wordt vele duizenden keren per seconde gemeten, en elke meting wordt weer als een setje bits vastgelegd. Zet je al die metingen achter elkaar af, dan hoor je weer muziek uit je speaker. Het bijzondere — en voor dit boek het belangrijkste — is dat het voor een computer, op het diepste niveau, geen enkel verschil maakt of die stroom nullen en enen een foto, een lied of een stuk tekst voorstelt. Het is allemaal precies hetzelfde soort ruw materiaal, alleen anders georganiseerd en anders geïnterpreteerd. Dat is ook precies waarom hetzelfde soort AI-technieken, zoals je in latere delen van dit boek zult zien, kan worden toegepast op zulke uiteenlopende zaken als foto’s, tekst en geluid: onder de motorkap is het allemaal patronen in reeksen bits.

De reden dat computers met deze simpele twee-waarden-schakelaars werken in plaats van met iets ingewikkelders, is puur praktisch: elektronische schakelaars die maar twee standen hoeven te onderscheiden (aan of uit) zijn extreem betrouwbaar, snel te maken, en goedkoop te produceren op enorme schaal. Zou je een schakelaar willen bouwen die tien verschillende standen nauwkeurig kan onderscheiden, dan wordt de techniek ineens veel lastiger en foutgevoeliger. Door slim gebruik te maken van heel veel simpele aan/uit-schakelaars in plaats van weinig ingewikkelde schakelaars, konden ingenieurs computers bouwen die zowel betrouwbaar als extreem snel zijn. Dat is de reden waarom, onder de motorkap, letterlijk alles in je telefoon, je laptop en elk ander digitaal apparaat is teruggebracht tot die eindeloze stroom nullen en enen.

Nu we weten waar computers hun instructies mee uitvoeren (bits), moeten we het hebben over wát die instructies eigenlijk zijn. Dat brengt ons bij het begrip programma, ook wel algoritme genoemd — een stappenplan dat een computer volgt om van een bepaalde invoer (input) tot een bepaalde uitvoer (output) te komen. Denk aan een algoritme als een recept: “neem twee eieren, klop ze los, voeg een snufje zout toe, verhit de pan, giet het mengsel erin, wacht twee minuten, keer het om.” Een computerprogramma is in essentie precies zoiets, alleen dan geschreven in een taal die de computer kan begrijpen (een zogeheten programmeertaal, die uiteindelijk wordt omgezet in die nullen en enen), en met stappen die veel preciezer en explicieter

moeten worden opgeschreven dan een menselijk recept. Een menselijk kok kan improviseren als het ei toch een beetje anders bakt dan verwacht; een computer kan dat niet, tenzij een programmeur van tevoren precies heeft voorzien wat er in zo'n situatie moet gebeuren en dat met zoveel woorden heeft opgeschreven. Cruciaal is: bij een traditioneel programma is het een mens die alle regels, alle stappen, expliciet en van tevoren uitschrijft. De computer voert vervolgens alleen maar uit wat die mens heeft opgeschreven — niets meer, niets minder.

Een ander vertrouwd voorbeeld maakt dit misschien nog concreter: de thermostaat aan je muur. Een simpele thermostaat volgt een piepklein, volledig door een mens bedacht algoritme: “meet de temperatuur; is die lager dan de ingestelde waarde, zet dan de verwarming aan; is die hoger of gelijk, zet de verwarming uit.” Geen enkel deel van dat proces vereist dat het apparaat iets “leert” — het volgt gewoon, feilloos en eindeloos herhaald, de regel die een ingenieur er ooit in heeft geprogrammeerd. Diezelfde logica, opgeschaald naar duizenden of miljoenen regels, ligt ten grondslag aan een enorm deel van de software die je dagelijks gebruikt: een rekenmachine-app, de spellingcontrole in een simpele teksteditor, het systeem dat een pinbetaling verwerkt, de kassa in de supermarkt. Al die programma's doen precies, en uitsluitend, wat een programmeur ervoor heeft opgeschreven. Ze zijn niet dom, in de zin dat ze feilloos hun werk doen binnen de grenzen van hun instructies, maar ze zijn ook niet flexibel: geef ze een situatie die de programmeur niet had voorzien, en ze lopen vast, geven een foutmelding, of doen iets duidelijk verkeerd.

Laten we dit heel concreet maken met een voorbeeld dat aansluit bij hoofdstuk 1: een spamfilter, gebouwd op de traditionele manier. Een programmeur zou dan met de hand een lijst met regels opstellen, zoiets als: “als het woord ‘gratis’ in de onderwerpregel staat, markeer de mail als verdacht”; “als de afzender een onbekend, willekeurig e-mailadres heeft met veel cijfers erin, verhoog het risico”; “als er drie of meer uitroeptekens achter elkaar staan, verhoog het risico verder”; “als het totale risico boven een bepaalde drempel komt, verplaats de mail naar de spammap.” Dit soort programma's kunnen prima werken voor de hand liggende gevallen. Het probleem is: de wereld zit vol uitzonderingen en verrassingen die een programmeur onmogelijk allemaal van tevoren kan bedenken. Wat als een legitieme winkel je een mail stuurt met “GRATIS VERZENDING” in de onderwerpregel? Wat als spammers hun tactiek veranderen en het woord “gratis” vermijden, maar wel handig gebruikmaken van andere trucs? Om elke nieuwe truc bij te houden, zou een programmeur voor altijd, dag in dag uit, met de hand nieuwe regels moeten blijven toevoegen — een eindeloze en uiteindelijk onhoudbare klus, want spammers passen zich sneller aan dan een team programmeurs kan bijbenen. Dit is precies het soort probleem waar de traditionele aanpak van programmeren tegen zijn grenzen aanloopt: er zijn simpelweg te veel uitzonderingen, te veel variatie, en te veel dingen die een mens niet allemaal kan voorzien en opschrijven.

En hier komt het cruciale omslagpunt waar dit hele hoofdstuk naartoe werkt — het verschil tussen traditioneel programmeren en AI, oftewel machine learning. Bij machine learning draait de aanpak volledig om. In plaats van dat een programmeur met de hand honderden regels opschrijft (“als dit, dan dat”), laat je de computer zelf de regels ontdekken door hem heel veel voorbeelden te laten zien. Bij een spamfilter zou dat er zo uitzien: je verzamelt tienduizenden, of zelfs miljoenen, e-mails die al zijn gemarkeerd als “spam” of “geen spam” (door mensen, in het verleden). Vervolgens laat je een machine-learning-systeem al die voorbeelden doorlopen, en de computer zoekt zelf, volledig automatisch, naar patronen die spam onderscheiden van normale mail — patronen die soms zo subtiel of talrijk zijn dat geen mens ze ooit allemaal met de hand zou kunnen opschrijven. Misschien ontdekt het systeem dat een combinatie van bepaalde woorden, in samenhang met de tijd waarop een mail wordt verstuurd en de structuur van de afzenderadres, een sterke aanwijzing is voor spam — een patroon dat te complex en te veelzijdig is voor een mens om expliciet in woorden te vangen, maar dat het systeem toch zelf, statistisch, weet te destilleren uit de enorme berg voorbeelden.

Dit verschil is zo fundamenteel dat het de moeite waard is om het nog eens met andere woorden samen te vatten, omdat het de sleutel is tot alles wat er in de rest van dit boek komt. Bij traditioneel programmeren is de volgorde: mens bedenkt de regels □ computer voert de regels uit op nieuwe gevallen. Bij machine learning is de volgorde compleet anders: mens verzamelt heel veel voorbeelden (met de juiste antwoorden erbij) □ computer ontdekt zelf de regels (of eigenlijk: de patronen) uit die voorbeelden □ die zelf-ontdekte patronen worden vervolgens toegepast op nieuwe, nog ongeziene gevallen. Het is een beetje het verschil tussen een kind leren fietsen door het exact te instrueren — “houd je linkerhand hier, trap met je rechtervoet zo hard, bij een bocht naar links leun je zo ver” — versus een kind gewoon laten fietsen, laten vallen, laten opstaan, en het zelf te laten aanvoelen wat werkt en wat niet, tot het na genoeg oefening vanzelf, zonder er nog bewust over na te denken, het evenwicht weet te bewaren. Bij die tweede aanpak kan niemand precies uitleggen welke exacte spierbewegingen het kind heeft “geleerd” — het kind kan het gewoon, na genoeg voorbeelden en oefening. Zo ook bij een AI-systeem: vaak kan zelfs de bouwer van het systeem niet meer precies aanwijzen welke exacte regel ervoor zorgt dat een bepaalde foto als “kat” wordt herkend. Het systeem heeft, na het zien van miljoenen voorbeelden, een enorm ingewikkeld web van patronen opgebouwd dat in de praktijk werkt, ook al is het voor mensen niet in simpele taal samen te vatten.

Dit brengt meteen een belangrijk inzicht met zich mee dat je door dit hele boek heen zult tegenkomen: AI-systemen zijn niet “geprogrammeerd” om te weten wat een kat is, in de zin dat iemand een lijst regels heeft opgeschreven over snorharen, oren en staarten. Ze zijn getraind — een woord dat je vanaf nu vaak zult tegenkomen — op grote hoeveelheden voorbeelden, en ze hebben daaruit zelf statistische patronen gedestilleerd die, gemiddeld genomen, goed genoeg werken om nieuwe, nooit eerder

geziene foto's correct te classificeren. Dat is meteen ook de verklaring voor waarom AI soms zo verrassend goed werkt in het ene geval, en zo verrassend slecht faalt in een ander, ogenschijnlijk vergelijkbaar geval: het systeem heeft geen echt begrip van "kathedraal" zoals jij dat hebt, het heeft alleen patronen geleerd uit de voorbeelden die het heeft gezien. Als die voorbeelden een bepaald soort situatie niet goed vertegenwoordigden, dan zal het systeem daar ook mee worstelen.

Dit brengt met zich mee dat machine learning, ondanks alle voordelen, ook een prijs heeft die traditioneel programmeren niet heeft: verlies van doorzichtigheid. Bij een traditioneel programma kun je, als je de code openklapt, in principe precies nagaan waarom het systeem een bepaalde beslissing nam — de regels staan er zwart-op-wit, stap voor stap, door een mens geschreven. Bij een getraind machine-learning-systeem is dat veel lastiger. Het "kennisbestand" van zo'n systeem bestaat niet uit leesbare regels, maar uit miljoenen of zelfs miljarden kleine, onderling verweven getallen die tijdens het trainingsproces zijn bijgesteld — een resultaat dat voor een mens vrijwel onmogelijk in gewone taal te ontleden is. Dat is de reden waarom onderzoekers weleens spreken van een "zwarte doos": je kunt zien wat erin gaat (de vraag, de foto) en wat eruit komt (het antwoord, de classificatie), maar het is verrassend lastig om precies te achterhalen waaróm het systeem tot die uitkomst kwam. Voor sommige toepassingen is dat onschuldig — het maakt niet zoveel uit waarom Netflix jou een bepaalde film aanraadt. Maar voor andere toepassingen, zoals wanneer een AI-systeem meebeslist over een medische diagnose, een lening, of een sollicitatie, is die ondoorzichtigheid een serieus punt van zorg, en het is een onderwerp waarop onderzoekers wereldwijd hard werken: hoe maak je zulke systemen beter uitlegbaar, zonder de kracht te verliezen die nu juist voortkomt uit het feit dat ze zelf, zonder handmatig geschreven regels, patronen kunnen ontdekken die geen mens had kunnen bedenken.

En daarmee land je automatisch bij de vraag die het onderwerp is van het volgende hoofdstuk: als een AI-systeem zijn "kennis" haalt uit voorbeelden in plaats van uit handgeschreven regels, dan is de kwaliteit en de hoeveelheid van die voorbeelden — die we voortaan data zullen noemen — allesbepalend voor hoe goed het systeem uiteindelijk werkt. Data is, zoals je in hoofdstuk 4 zult lezen, de brandstof waarop de hele motor van moderne AI draait. Zonder goede, rijke, betrouwbare data blijft zelfs de knapste technische architectuur nutteloos — net zoals de beste automotor ter wereld niets waard is zonder brandstof om op te rijden.

Wat je nu weet

- Een computer voert razendsnel en precies instructies uit, zonder enig 'begrip' van wat hij doet.

- Traditioneel programmeren (vaste regels) loopt vast op de eindeloze uitzonderingen van de echte wereld — machine learning draait dit om door de computer zelf patronen te laten ontdekken in voorbeelden.
- Die omslag — van handgeschreven regels naar geleerde patronen — is het scharnierpunt van de rest van het boek.

Hoofdstuk 4: Data — de brandstof van AI

In het vorige hoofdstuk kwam een woord voorbij dat vanaf nu een van de belangrijkste begrippen in dit boek zal zijn: data. Het is tijd om daar goed bij stil te staan, want zonder data — en zonder te begrijpen wat data precies is en waarom het zo cruciaal is — kun je onmogelijk snappen hoe moderne AI werkt. Een veelgehoorde uitspraak onder mensen die met AI werken is dat data de nieuwe brandstof is, of soms zelfs “het nieuwe goud”. Dat klinkt misschien als een cliché uit een technologieartikel, maar de vergelijking klopt verrassend goed, en in dit hoofdstuk gaan we uitleggen waarom.

Laten we eerst helder krijgen wat “data” nu eigenlijk precies betekent, want het is een woord dat we tegenwoordig overal horen zonder dat het altijd wordt uitgelegd. Data is, in de simpelste bewoordingen, alles wat je kunt vastleggen en bewaren. Dat klinkt misschien vaag, dus laten we het concreet maken: een foto die je met je telefoon maakt, is data. De tekst van een boek, een e-mail, of een appje, is data. Een geluidsopname van iemand die praat, is data. De cijfers in een boekhoudprogramma — hoeveel iets kost, hoeveel er verkocht is — zijn data. Zelfs iets ogenschijnlijk onzichtbaars als jouw klikgedrag op een website — welke knoppen je aanklikt, hoe lang je naar een bepaalde pagina kijkt voordat je doorscrollt — wordt vastgelegd en is dus ook data. Kortom: elk stukje informatie dat ergens digitaal is opgeslagen, in welke vorm dan ook — tekst, beeld, geluid, getallen, gedrag — valt onder de noemer data. Het is de grondstof, het ruwe materiaal, waarmee AI-systemen worden gebouwd, op precies dezelfde manier waarop meel, water en gist de grondstoffen zijn waarmee een bakker brood bakt.

Binnen die brede categorie data maken onderzoekers ook nog onderscheid tussen wat ze gestructureerde en ongestructureerde data noemen, en dat onderscheid is handig om te kennen omdat het laat zien hoe verschillend data eruit kan zien.

Gestructureerde data is netjes georganiseerd in vaste velden, zoals een spreadsheet met rijen en kolommen: naam, leeftijd, postcode, aankoopbedrag. Een computer heeft daar van oudsher weinig moeite mee, want elk gegeven staat op een voorspelbare plek. Ongestructureerde data is veel rommeliger van vorm: een vrij geschreven e-mail, een foto, een geluidsopname, een video. Er staat geen keurig hokje omheen dat zegt “dit stuk is de naam” of “dit stuk is het bedrag” — het is gewoon een grote, vormeloze brei aan informatie. Het grootste deel van alle data die er in de wereld bestaat, is ongestructureerd: denk aan alle foto’s op smartphones, alle appjes, alle YouTube-video’s. Lange tijd konden computers relatief weinig met dat soort rommelige data beginnen, en juist een van de grote prestaties van moderne AI is dat het, dankzij technieken als deep learning waar je in deel 2 meer over leert, eindelijk goed grip heeft gekregen op precies dit soort ongestructureerde, rommelige, mensachtige data.