

Inhoud

Voorwoord	9
Prof. Tom Braekeleirs (Digitale Medische Technologie, Faculteit Geneeskunde Universiteit Gent)	
Inleiding	13
Prof. Wim Van Biesen (Diensthoofd Nefrologie UZ Gent, Vakgroep inwendige ziekten en pediatrie Universiteit Gent) en Prof. Sigrid Sterckx (Universiteit Gent, Moraalfilosofie, bio-ethiek en politieke filosofie)	
DE STAND VAN ZAKEN IN AI- TECHNOLOGIE	19
Hoofdstuk 1. Een heldere kijk op artificiële intelligentie, machinelearning en <i>Large Language Models</i>	21
Prof. Joni Dambre (Universiteit Gent, Faculteit Toegepaste Wetenschappen, Internet Technology and Data Science Lab – ID Lab)	
INTERLUDIUM. Hulde aan de bijdrage van vrouwen tot computerwetenschap en AI: Ada Lovelace (1815-1852)	49
Prof. Sigrid Sterckx (Universiteit Gent, Moraalfilosofie, bio-ethiek en politieke filosofie)	
DEEL I VERSCHILLENDE KENNISTAKEN IN VERSCHILLENDE STADIA VAN DE GENEESKUNDIGE BESLUITVORMING	57
Hoofdstuk 2. Predictie en theoretische modellen: De beloften en beperkingen van big data	59
Prof. Wim Van Biesen (Diensthoofd Nefrologie UZGent, vakgroep Inwendige ziekten en pediatrie Universiteit Gent)	

Hoofdstuk 3. AI en nieuwe vormen van klinische studies: Gladjes naar evidentie of evidentie op glad ijs? 85

Dr. Johan Steen (datawetenschapper – postdoctoraal onderzoeker UZGent) en Prof. Stijn Van Steelandt (Universiteit Gent, Vakgroep toegepaste wiskunde en statistiek)

Hoofdstuk 4. Het unieke karakter van medische besluitvorming: Implicaties voor het gebruik van AI in geneeskunde 115

Prof. Wim Van Biesen (Diensthoofd Nefrologie UZGent, vakgroep inwendige ziekten en pediatrie Universiteit Gent) en Prof. Sigrid Sterckx (Universiteit Gent, Moraalfilosofie, bio-ethiek en politieke filosofie)

DEEL II VOORBEELDEN OP DE WERKVLOER 135

Hoofdstuk 5. *Omics everywhere*: AI en big data in de diagnostische geneeskunde 137

Prof. Bruno Verhasselt (Hoofd Labodiagnostiek UZGent) en Prof. Bruce Poppe (Diensthoofd Genetica UZGent)

Hoofdstuk 6. Geneesmiddelenontwikkeling en AI: Rollen de moleculen nu gewoon uit de computer? 161

Prof. Evelien Wynendaele, Dra. Hannah Hirmz en Prof. Bart De Spiegeleer (Laboratory of Drug Quality & Registration, Faculteit Farmaceutische Wetenschappen Universiteit Gent)

Hoofdstuk 7. *Clinical Decision Support Systems*: Algoritmes die ondersteunen versus algoritmes die sturen 181

Prof. Kirsten Colpaert (Hoofd Data Science Institute UZGent) en Prof. Johan Decruyenaere (Voorzitter Klinisch InnovatieFonds UZGent)

Hoofdstuk 8. *Medicine meets Extended Reality*: Over chirurgen, 3D-printers en andere verhalen 211

Dr. Renaat Coopman (Dienst plastische heelkunde UZGent en Cancer Research Institute Ghent) en Prof. Nikdokht Rashidian (Dienst Hepatobiliaire Heelkunde UZGent)

DEEL III BREDERE VRAGEN VOOR ZORGVERLENERS, PATIËNTEN EN DE MAATSCHAPPIJ	237
---	------------

Hoofdstuk 9. Gedeelde besluitvorming in tijden van AI: Een <i>ménage à trois</i>?	239
--	------------

Dr. Edgard Eeckman (Voorzitter vzw PatientEmpowerment) en Prof. Wim Van Biesen (Diensthoofd Nefrologie UZGent, Vakgroep inwendige ziekten en pediatrie Universiteit Gent)

Hoofdstuk 10. AI in gezondheidszorg: Wie is aansprakelijk wanneer het fout loopt?	261
--	------------

Prof. Kristof Vanassche (Onderzoeksprofessor Gezondheidsrecht Faculteit Rechten Universiteit Antwerpen) en Prof. Sigrid Sterckx (Universiteit Gent, Moraalfilosofie, bio-ethiek en politieke filosofie)

Hoofdstuk 11. Naar een totale automatisering van de geneeskunde?	291
---	------------

Prof. Sigrid Sterckx (Universiteit Gent, Moraalfilosofie, bio-ethiek en politieke filosofie) en Dr. Els Dufraimont (Dienst Gynaecologie, Imelda Ziekenhuis)

Hoofdstuk 12. AI als transformator van geneeskunde: Een pleidooi voor gezamenlijke betrokkenheid	319
---	------------

Prof. Wim Van Biesen (Diensthoofd Nefrologie UZGent, vakgroep inwendige ziekten en pediatrie Universiteit Gent) en Prof. Sigrid Sterckx (Universiteit Gent, Moraalfilosofie, bio-ethiek en politieke filosofie)

Voorwoord

Prof. Tom Braekeleirs

AI is een blijvertje. Ziezo... Open deur... Ingetrapt.

De komst van ChatGPT eind november 2022 confronteerde de wereld met een revolutionair technologisch hoogstandje op een schaal die we nooit eerder zagen. Eén van de vorige grote revoluties, het internet, sloop met mondjesmaat onze levens binnen – want om goed te kunnen snappen wat het was, had je iets nodig... het internet. Maar nu een groot deel van de wereld over internet beschikt, werd het een makkelijke uitrol. En de cijfers logen er niet om. Binnen vijf dagen na de lancering hadden meer dan een miljoen mensen geëxperimenteerd met de tool. Twee maanden later was dat al gestegen tot honderd miljoen. In het begin – laten we wel wezen – gebruikten mensen de tool vooral om grapjes te genereren, maar intussen begint de maturiteitsfase en zien we reële scenario's ons dagelijks leven binnensluipen.

Dankzij de komst van ChatGPT gingen de spotlights niet alleen naar generatieve AI (de mogelijkheid om nieuwe content te genereren zoals tekst, beelden, audio...). Ze brengt AI in zijn bredere context uit een winterslaap waar die de voorbije jaren een beetje was in verzeild geraakt. Niettemin is AI geen nieuw begrip. Al sinds de opkomst van machines droomt de mens over zelfdenkende technologie. Zeventig jaar geleden werd de term 'AI' (of voor taalpuristen 'KI' – kunstmatige intelligentie) voor het eerst geponeerd. Initieel als een bijna filosofische denkoefening, nog vóór de uitvinding van de personal computer. In een tijdperk waarin analoge machines de verwachtingen prikkelden, ontstonden dromen over wat AI ooit zou kunnen worden. De weg ernaartoe vergde decennia: van eenvoudige, regelgebaseerde systemen naar expertsystemen tot nu, met de opkomst van de generatieve AI.

Deze keer is de sprong voorwaarts zo groot dat we die moeilijk kunnen verwaarlozen. Uiteraard is AI veel meer dan alleen de generatieve AI, die nu vaak

de voorgrond inneemt. De wereld van machine learning en neurale netwerken zijn zo mogelijk nog interessanter en hoopvoller. Dat wil niet zeggen dat nu alle problemen die de mensheid ooit heeft ervaren als sneeuw voor de zon zullen verdwijnen, wél dat we optimistisch mogen zijn dat een en ander begint te verschuiven. Zo ook in de gezondheidszorg. Dat is een sector die zo belangrijk is in ons dagelijks leven dat iedereen er zich wel iets bij kan voorstellen. Het is er een die vraagt om een uiterst zorgvuldige, doordachte en verantwoorde omgang met deze nieuwe technologieën. Dat begint vaak met een goed begrip van wat het is, wat het kan en hoe we het op de juiste manier kunnen toepassen.

En daar draagt dit boek heel wat aan bij. Een beter begrip van de achterliggende werking van AI-systemen neemt hopelijk heel wat angst en weerstand weg. Want die is er helaas ook. Aangewakkerd door het vaak ongebreidelde enthousiasme van de ontwikkelaars van technologie, als ze met veel poeha en bravoure voor de zoveelste keer aankondigen dat <vul hier een productnaam in> de wereld écht tot een betere plek gaat maken. Telkens opnieuw presenteren ze de nieuwe technologie dan als uniek, revolutionair en wereldschokkend... Maar de realiteit toont soms minder optimistische resultaten, zoals ook dit boek met een aantal voorbeelden illustreert.

Toch mogen we niet concluderen dat het nog te vroeg is voor AI. Juist door te leren wat AI allemaal kan, maken we vooruitgang. Door te leren waar we het kunnen toepassen, te blijven zoeken naar scenario's waar het wel steek houdt en te leren uit de fouten die onderweg gemaakt worden. Voortschrijdend inzicht, noemen we dat, en dat bouw je alleen op door technologie te omarmen. In tijden van AI betekent dit dat we, letterlijk, samen moeten leren.

In mijn lessen nodig ik studenten uit om kritisch te kijken naar de mogelijkheden en de resultaten. Maar ook kritiek moet je in een context van optimisme plaatsen. Constructieve kritiek, zo je wil. Dit boek is een uitnodiging om mee te denken, binnen een constructief-kritische context.

Je kan je afvragen hoe lang deze fase gaat duren. Dat is een moeilijke vraag en een glazen bol heb ik helaas niet in mijn bezit. De exponentiële groei van de mogelijkheden van generatieve AI in de afgelopen jaren biedt hoop, maar zoals de geschiedenis laat zien, komt er na een fase van groot enthousiasme een dal van ontgoocheling voordat de werkelijke meerwaarde zichtbaar wordt. Die meerwaarde hangt altijd af van de context waarin AI wordt toegepast en is daardoor moeilijk te veralgemenen. Een aantal scenario's zijn *quick-wins* voor AI, zoals administratieve ondersteuning, *clinical decision support* systemen of door

gigantisch grote datasets struinen. Dit boek levert een reeks van voorbeelden binnen verschillende vakgebieden van de geneeskunde, zonder voorbij te gaan aan juridische en ethische denkkaders en overpeinzingen.

Want er zijn ook valkuilen waarvan we ons bewust moeten zijn. *Deskilling* (het verlies van competenties omdat we systemen en tools hebben die (deels) het werk voor ons doen) is binnen dit domein een belangrijke component om in de gaten te houden. Want zelfs de meest accurate AI-systemen willen er wel eens naast zitten. En wij – als mens – zijn nog steeds diegene die tegenover de patiënt zitten, en hopelijk blijft dat nog lang zo. De angst dat AI en andere digitale tools de zorg ontmenselijken, is een betoog waar ik niet in meega. Ik ben juist overtuigd van het tegenovergestelde. Als we technologie correct inzetten – binnen de juiste denkkaders, met de juiste juridische achtergrond, en niet alleen om efficiëntieverhoging na te streven – zal dit hopelijk weer meer ruimte geven om de mens in de zorg meer op de voorgrond te krijgen, weg van een computerscherm, dat letterlijk en figuurlijk vaak een barrière vormt in de patiënt-arts-relatie. En vermoedelijk zal er een moment komen waarop de mens en de machine zo'n sterke tandem vormen, dat AI binnen het kader van *Evidence Based Medicine* onmisbaar wordt en het bijna patiëntonvriendelijk zou zijn om het niet te gebruiken. Voorlopig zijn we daar nog niet. En in tijden van AI, moet je altijd voorzichtig zijn met uitspraken. Vandaar de klemtoon op 'nóg niet'.

Hoe het zal evolueren, weet ik ook niet. Maar dat het een impact heeft en dat we de discussie moeten voeren, is meer dan zeker. Tijd voor een optimistisch en kritisch constructief debat. En laat dit boek daar een goede leidraad in zijn. Want AI niet omarmen, zou mijns inziens een denkfout zijn. De potentiële voordelen zijn te groot om het niet te proberen. Want dan gooien we het kind met het badwater weg. Ziezo... Nog een open deur... Ingetrapt.

Inleiding

Over de beloftevolle toekomst van artificiële intelligentie (AI) in de gezondheidszorg is al veel inkt gevloeid. Zoals altijd bij wetenschappelijke innovatie zijn er voor- en tegenstanders in alle gradaties: van profetische *diehard believers*, over gematigde enthousiastelingen of twijfelaars, tot rabiate tegenstanders en doemdenkers. Er wordt geschreven dat sommige medische specialismen zullen verdwijnen met de opkomst van AI, waarbij vooral de radiologen in het oog van de storm lijken te zitten. Zoals bij alle nieuwe technologie is echter een wetenschappelijke houding aangewezen. Proberen te begrijpen hoe het werkt, wat de resultaten zijn in de dagelijkse praktijk van zorgverleners, en rigoureus evalueren of deze uitkomsten consistent zijn of dat het om toevalstreffers gaat. Naast het verbeteren van de technologie zelf is het dus ook noodzakelijk om de impact van de invoering ervan in de unieke context van de geneeskunde goed te bestuderen.

Alle gebruikers van AI-technologie, dat zijn zowel artsen als andere zorgverleners maar ook het ziekenhuismanagement en beleidsmakers, moeten geïngereerde en realistische verwachtingen hebben over wat de technologie kan en niet kan. Als we onze gezondheidszorg echt willen verbeteren, moeten we AI-technologie die een duidelijke meerwaarde heeft invoeren. En als ze geen meerwaarde heeft, of juist meer problemen of valkuilen introduceert, moeten we ze vermijden. Doen we dat niet, dan zal de implementatie van AI uitlopen op suboptimale uitkomsten en een nieuwe golf van AI-cynisme. We zouden dan kansen verloren laten gaan om AI te ontwikkelen en te gebruiken in toepassingen waarbij ze wel een meerwaarde kan creëren.

Om dit te vermijden, en dus een optimaal gebruik van AI te bewerkstelligen, is het uiterst belangrijk dat iedereen die betrokken is bij of bezorgd over geneeskunde en gezondheidszorg, van artsen en patiënten tot management en beleidsmakers, de kennis en vaardigheden verwerft om mee na te denken over AI-technologie en hoe ze op een verstandige, waardevolle en rechtvaardige manier ingezet kan worden. In de gedigitaliseerde maatschappij waarin we tegenwoordig leven, gaat dit ons *allemaal* aan.

Fast forward naar generatieve AI zoals ChatGPT. Zeker sinds er ook een versie van bestaat met een stem die zegt: 'Ik ben er om je te helpen en om je gezelschap te houden', is het debat over wat AI-technologie allemaal 'kan' en wat ze op basis daarvan 'is', nog intenser geworden. Wanneer het gaat over een context zoals de geneeskunde, waarin levens op het spel staan, is het dus ook belangrijk dat we ons afvragen hoe zorgverleners, beleidsmakers en patiënten ernaar kijken. Bijna alle mensen hebben immers sterk de neiging om technologie te 'antropomorfiseren', wat betekent dat we er *menselijke eigenschappen aan toeschrijven*. Heel wat mensen zijn ervan overtuigd dat generatieve AI allerlei dingen kan die je normaal met een mens zou associëren, maar niet met machines of robots. De vraag of er dan sprake is van een overgang van 'machine' naar 'persoon' is een heel boeiende kwestie, en eigenlijk heel relevant voor het thema van dit boek: het gebruik van AI in de geneeskunde.

Stel je even voor dat je een handige timmerman bent, die een bedrijfje leidt waar houten meubels worden gefabriceerd. Een tiental arbeiders gaat dagelijks aan de slag met hout, leer, stof en allerlei soorten kunststof om die comfortabele en kwalitatief hoogwaardige zitbanken te maken waar jullie bedrijf zo beroemd om is. De verkoop loopt als een trein, en je wilt de productiviteit verhogen door te automatiseren. Met andere woorden, je bent van plan om bepaalde handelingen, de controle en de planning die nu nog worden uitgevoerd door je werknemers of door jezelf, te vervangen door automatische hulpmiddelen of machines. Je wilt daarbij echter niet inboeten aan kwaliteit, en je beslist om de bestaande kennis en expertise van je medewerkers in te bouwen in de nieuwe geautomatiseerde productielijn. Je moet dus eerst goed in kaart brengen *hoe en waarom* je medewerkers bepaalde handelingen uitvoeren voor je deze kunt omzetten in een geautomatiseerd proces.

Ook als we AI willen inzetten om de besluitvorming in onze gezondheidszorg te verbeteren, moeten we eerst goed begrijpen hoe het proces van *kennisverwerving* precies verloopt binnen de gezondheidszorg, en welke stappen zorgverleners eigenlijk zetten bij hun *besluitvorming*.

In het eerste hoofdstuk krijg je daarom een boeiend overzicht van de manier waarop AI precies werkt, en welke verschillende systemen er bestaan. Ook als je nog geen voorkennis hebt, ben je met dit hoofdstuk bij, niet alleen met AI-technologie die al lang wordt gebruikt, maar ook the *new kids on the block*, zoals generatieve AI en de zogenaamde *Large Language Models*. Al deze benaderingen zijn immers ook in de geneeskunde relevant.

In de daaropvolgende drie hoofdstukken gaan we dieper in op de zogenaamde *big data*, want die spelen een cruciale rol bij de ontwikkeling van AI. We willen vooral begrijpen hoe big data werken, en vanuit dat inzicht al nagaan wat we er wel en niet van kunnen *verwachten*. We bekijken eerst hoe we kennis kunnen opdoen over de wereld. Dit boek focust natuurlijk op de wereld van de gezondheid en geneeskunde, dus moeten we vooral nagaan hoe kennisverwerving en besluitvorming daar anders zijn dan wanneer het gaat over andere facetten van onze wereld (bijvoorbeeld het maken van comfortabele zitbanken).

Vervolgens illustreren we in Deel II in vier verschillende praktijkgerichte hoofdstukken met behulp van heel concrete voorbeelden hoe *big data* en AI ons kunnen helpen om sommige taken in de geneeskunde beter, sneller en efficiënter uit te voeren. Op basis van de verhalen van mensen die er op dagelijkse basis mee werken in ziekenhuizen, krijg je een inkijk in de manieren waarop AI wordt ingezet in klinische *laboratoria*, onder andere om bloed en weefsels te onderzoeken en voorspellingen te doen over erfelijke risico's op bepaalde aandoeningen. Je ontdekt bovendien dat AI nu al een grote rol speelt in het domein van de *geneesmiddelen*. Dat geldt zowel voor het ontdekken van nieuwe manieren waarop geneesmiddelen die al langer bestaan kunnen worden gebruikt, als voor het ontwikkelen van nieuwe geneesmiddelen. Nog een andere praktijkcontext waarover de experts die je in dit boek binnenlaten in hun klinische praktijk op een heldere manier hun fascinatie delen, is het gebruik van AI om een soort van gps te ontwikkelen die artsen en andere zorgverleners kan toelaten om het zorgtraject van een patiënt te navigeren. Kortom, de hulp van computers om complexe *medische beslissingen* te kunnen nemen over een patiënt of een groep van patiënten. Vervolgens word je meegenomen naar de fascinerende wereld van de *chirurgie* en zie je hoezeer AI daar al ingeburgerd is via het gebruik van chirurgische robots, maar ook hoe AI in geavanceerde ziekenhuizen wordt gebruikt bij allerlei fases van de *voorbereiding* van een operatie, het vele werk dat sommige chirurgen samen met computerwetenschappers verrichten om specifieke ingrepen tot in de puntjes te plannen, lang voordat de patiënt de operatiezaal wordt binnengebracht.

In Deel III, het laatste deel, spitsen we ons vooral toe op de *maatschappelijke* weerslag van deze nieuwe gedigitaliseerde aanpak van de geneeskunde. Het staat immers als een paal boven water dat de overschakeling naar meer digitalisering, en dus het gebruik van big data en AI, een enorme invloed zal hebben op hoe we in de heel nabije toekomst gezondheidszorg *krijgen* en *ervaren*, en waarschijnlijk ook op hoe we naar 'ziekte' en 'gezondheid' kijken en zelfs op wat we onder die termen verstaan. Als we ons bijvoorbeeld afvragen wat de term *arts*

eigenlijk betekent, of beter, wat die zou kunnen of moeten betekenen, dan zou een mogelijk antwoord kunnen zijn: een expert die jij als patiënt vertrouwt om je te helpen bij het maken van belangrijke maar moeilijke keuzes in omstandigheden van stress en kwetsbaarheid. Als dat een aannemelijk antwoord is, moeten we ons dus de vraag stellen of een AI-systeem ooit de plaats zou kunnen innemen van een arts zonder dat wezenlijke aspecten van het beroep of de rol van arts verloren gaan.

Bij een AI-systeem ontstaat de ‘intelligentie’ *from scratch* via de data die het systeem ontvangt. Zorgverleners ontvangen en verwerken echter veel meer data dan alleen de digitaliseerbare feiten betreffende hun patiënt, denk maar aan lichaamstaal, of sociale context, emoties, voorgeschiedenis en sociale relaties. Op die manier ontstaat een onnoemelijk veel rijker beeld van de patiënt als ‘mens in een context’. Er blijft de noodzaak aan evenwicht, waarbij zowel de mens-zorgverlener als de AI-technologie een bijdrage kunnen leveren. De *interpretatie* van een CT-scan of van een weefselstaal staat immers los van de sociale achtergrond: een melanoom is een melanoom. Een patiënt is echter meer dan een ziekte, en een ziekte is meer dan een diagnose. Wat met begrijpen, betekenis geven, voelen, zingeving zoeken? Niet alleen door patiënten trouwens, maar ook door zorgverleners. Wat met angst voor sterfelijkheid? Niet alle vraagstukken in de gezondheidszorg kun je oplossen of laten verdwijnen door er meer ‘data’ tegenaan te gooien. Als we het hebben over de plaats van AI in de gezondheidszorg, moeten we dus in de eerste plaats goed nadenken over wat volgens ons eigenlijk het *doel* is van gezondheidszorg. Wat willen we bereiken? Waar willen we naartoe? Is er een gedeelde visie op zorg en op mensen, een visie die als grondslag en inspiratiebron dient voor het ‘beleid’? Pas dan kan worden gekeken of en hoe AI kan helpen.

Hoe kunnen *zorgverleners* een evenwicht zoeken tussen *enerzijds* het inzetten van AI als ze een zeer welkome ondersteuning kan bieden en soms zelfs compleet nieuwe soorten zorg mogelijk kan maken om noden te lenigen waar tevoren weinig of niets kon voor worden gedaan, en *anderzijds* de vraag te blijven stellen of zijzelf geen dingen uit het oog verliezen die nog altijd heel belangrijk zijn, maar die in een gedigitaliseerd zorgsysteem van de radar kunnen verdwijnen omdat ze niet in data kunnen worden uitgedrukt of beschreven? En wat met de implicaties voor *patiënten*? Kan AI helpen om hen meer te betrekken bij beslissingen over hun zorg? Waarop moeten we letten om te voorkomen dat mensen uitgesloten worden of uit de boot vallen? Wij als maatschappij moeten immers enorm aan digitale geletterdheid en AI-geletterdheid werken, want louter het feit dat iemand een smartphone bezit, betekent nog niet dat hij digitaal

voldoende mee is om zijn rechten als patiënt effectief uit te oefenen. Bovendien worden niet alleen patiënten maar ook zorgverleners zelf geconfronteerd met een golf van nieuwigheden en tools die ze moeten gebruiken, zonder dat ze in hun opleiding geleerd hebben hoe die werken en wat de meerwaarde en de risico's ervan kunnen zijn.

De vraag wie aansprakelijk is – dus letterlijk: wie er moet worden aangesproken – wanneer er iets fout loopt met AI-technologie in de gezondheidszorg, is ook relevant, zowel voor zorgverleners als voor gezondheidsinstellingen en patiënten. Het antwoord is ook sterk bepalend voor hoe AI zal worden ingezet, en hoe ermee omgegaan zal worden. Dit is geen gemakkelijke vraag en het antwoord hangt af van veel factoren, maar in Europa worden alvast grote stappen gezet om patiënten die schade lijden als gevolg van het gebruik van AI-technologie beter te beschermen.

Net zoals voor elke andere technologie is het noodzakelijk om samen na te denken over hoe we AI- en big data-technologie het best kunnen en willen inzetten om zoveel mogelijk maatschappelijke meerwaarde te creëren. Geen belangrijkere en meer dringende context om dit te doen dan de geneeskunde, vonden wij! Zo is het plan ontstaan voor het boek dat je in handen hebt. Laat je op sleeptouw nemen door de vele artsen, onderzoekers en andere ervaringsdeskundigen die niet alleen hun kennis en rijke ervaringen, maar ook hun hoop, verwachtingen, toekomstdromen, bezorgdheden en waarschuwingen delen in dit boek. *Enjoy the ride...* en minstens even belangrijk: *join the conversation!*

Sigrid Sterckx en Wim Van Biesen

De stand van zaken in ai-technologie

HOOFDSTUK 1.

Een heldere kijk op artificiële intelligentie, machinelearning en *Large Language Models*

Joni Dambre

“We’ll show you that you can build a mind from many little parts, each mindless by itself.”

(Marvin Minsky, een van de grondleggers van AI, 1986)

EEN BEETJE GESCHIEDENIS

De ‘A’ van kunstmatig

In de context van AI betekent artificieel (*artificial* in het Engels) hetzelfde als kunstmatig, in tegenstelling tot *natuurlijk*, of *biologisch*. Kunstmatige intelligentie verwijst met andere woorden naar intelligentie of intelligent gedrag van een (door mensen ontworpen) machine.

Al eeuwen dromen en filosoferen mensen over intelligente machines, maar pas in de eerste helft van de 20ste eeuw was er voldoende kennis om deze dromen ook te beginnen te realiseren. Na de industriële revolutie hadden door motoren aangedreven machines een belangrijke plaats veroverd in fabrieken, om het productieproces van goederen sneller en goedkoper te laten verlopen. Vanaf het eind van de 19de eeuw veroverden verschillende soorten machines, zoals de automobiel en de telefoon, het dagelijks leven. Al snel werd duidelijk dat deze uitvindingen ook militair enorm nuttig waren. In de marge van de Eerste Wereldoorlog kreeg de ontwikkeling van machines voor wapentuig, transport en communicatie dan ook een enorme *boost*.

De automatisering van berekeningen werd het volgende doel. Al in de loop van de 19de eeuw werden verschillende *mechanische* machines ontworpen waarmee bepaalde berekeningen automatisch konden worden uitgevoerd: de eerste *rekenmachines* dus. In de eerste helft van de 20ste eeuw begon men eenvoudige

digitale rekensystemen te ontwerpen met *elektrische* in plaats van mechanische componenten. In de Tweede Wereldoorlog kreeg dit soort onderzoek (alweer) een extra boost. Zo was het automatisch ontcijferen van de geheime codes van de vijand een van de hoofddoelen van het door de Britse overheid gefinancierde werk van de Britse wiskundige *Alan Turing*. Hiermee zou men immers het werk van *menselijke* rekenaars (letterlijk *computers* in het Engels) kunnen versnellen. In 1944 werd de eerste *Colossus*-computer in gebruik genomen. Door schakelaars te verzetten, kon men kiezen welke berekening zou worden uitgevoerd. Dit was dan ook de eerste *programmeerbare* elektronische computer.

De *Colossus* was speciaal ontwikkeld om één specifieke stap in het ontcijferen van gecodeerde Duitse berichten te automatiseren. In de jaren dertig stelde Alan Turing dat het beter zou zijn om een *algemeen bruikbare computer* (een zogenaamde *general purpose computer*) te ontwikkelen, waarmee je heel veel verschillende soorten taken zou kunnen uitvoeren door de computer op een gepaste manier te *programmeren*. Op dat moment had hij echter nog geen concreet idee van hoe een dergelijk apparaat eruit zou moeten zien in de praktijk. Pas aan het einde van de Tweede Wereldoorlog werd de eerste stap gezet naar de praktische realisatie van deze ideeën toen het Amerikaanse leger startte met de bouw van de ENIAC (*Electronic Numerical Integrator And Computer* - 1946). Deze eerste algemeen bruikbare computer moest men wel nog altijd 'programmeren' door manueel verbindingen tussen de componenten aan te passen, een arbeidsintensief werkje dat werd uitgevoerd door zes vrouwelijke programmeurs. Voor de opvolger van deze computer, de EDVAC (*Electronic Discrete Variable Automatic Computer* - 1949), bedacht de Hongaars-Amerikaanse wiskundige en fysicus *John von Neumann* hoe een programma, in de vorm van een reeks instructies, kon worden opgeladen in het geheugen van de computer. Hierdoor werden de instructies eigenlijk een deel van de invoer voor de uit te voeren berekeningen, zodat je op eenzelfde computer ontelbare programma's kunt uitvoeren zonder dat je (manueel) structurele aanpassingen moet doen aan de hardware. Het werkingsprincipe van deze zogenaamde *von Neumann-architectuur* vormt nog altijd de basis van de meeste digitale computers.

De eerste automatische rekeneenheden en programmeerbare computers maakten gebruik van radiobuizen en andere vrij grote componenten. Hierdoor werden ze al snel heel groot en heel duur. De bouw van de ENIAC, bijvoorbeeld, kostte destijds al bijna 500.000 dollar en het toestel woog 30 ton. Pas in 1954 kwamen de eerste transistors op de markt, waarmee je veel kleinere computers kon bouwen. In 1955 kon ook het brede publiek voor het eerst een digitaal rekenproduct op basis van transistors kopen: de IBM608 rekenmachine.

Pas in de jaren zestig was de technologie van *geïntegreerde schakelingen*, waarmee je heel veel transistors samen op één chip kunt zetten, voldoende ver ontwikkeld om er computers mee te kunnen bouwen. De massaproductie van computers voor het brede publiek (*personal computers*) zou pas echt beginnen aan het eind van de jaren zeventig. Doordat men door de jaren heen steeds meer en steeds kleinere transistors kon samenbrengen op eenzelfde chip zijn de hedendaagse computers al miljarden keren krachtiger dan hun eerste commerciële voorouders. Zonder deze gigantische digitale rekenkracht zou het onmogelijk zijn om hedendaagse AI-systemen te ontwikkelen of te gebruiken.

De 'I' van intelligentie

Op het vlak van nauwkeurig en betrouwbaar rekenen zijn digitale computers veel sterker dan mensen, en zolang de hardware blijft werken, is hun geheugen perfect. Maar ze volgen, stap voor stap en uiterst nauwkeurig, de instructies in door mensen geschreven programma's. Dat was nog lichtjaren verwijderd van waar de vroege dromen over intelligente machines over gingen: machines die met mensen kunnen communiceren, die kunnen bijleren, die zelfstandig slimme strategieën kunnen ontwikkelen in nieuwe situaties en weloverwogen beslissingen kunnen nemen.

Een algemeen aanvaarde definitie van wat als 'intelligentie' beschouwd kan worden is er niet, maar dit begrip werd wel lange tijd in hoofdzaak afgemeten aan menselijke eigenschappen. Op basis van kennis uit verschillende wetenschapsdomeinen zoals biologie, psychologie, wiskunde en statistiek richtten onderzoekers in de eerste helft van de 20ste eeuw zich op het proberen te reproduceren van heel verschillende uitingen van intelligentie, zoals het bewijzen van wiskundige stellingen, het zoeken van een uitweg uit een doolhof, het oplossen van een puzzel, het winnen van een 'moeilijk' spel zoals schaken, het visueel herkennen van objecten, of het beantwoorden van vragen die gesteld werden via geschreven tekst.

De eerste, heel diverse ideeën en experimenten rond 'intelligente machines' ontstonden lang vóór er programmeerbare computers beschikbaar waren. Vaak waren die geïnspireerd door ontdekkingen en hypothesen over de werking van het brein en de veronderstelde mechanismen achter het *leren* van dieren en mensen. Zo waren onder andere de prachtige tekeningen en gedetailleerde beschrijvingen van de structuur van biologische neuronen door Ramón y Cajal de inspiratie voor McCulloch en Pitts om in 1943 een eenvoudig wiskundig model op te stellen voor de werking ervan. Dit vormde later de belangrijkste bouwsteen

van *artificiële neurale netwerken* die vandaag de meest gebruikte AI-techniek zijn. De fascinatie voor het menselijk brein bleef een bron van inspiratie. In 1952 stelde Marvin Minsky zijn SNARC (*Stochastic Neural Analog Reinforcement Calculator - 1951*) voor, de eerste machine die een vorm van kunstmatige intelligentie realiseerde. Het werkingsmechanisme van de SNARC was gebaseerd op een 'brein' van 40 kunstmatige neuronen die konden leren hoe je uit een doolhof ontsnapt. Dit leren was gebaseerd op de theorieën van Donald Hebb enerzijds, die stelde dat de sterkte van verbindingen tussen neuronen wordt aangepast op basis van correlaties tussen neuronactiviteit, en Burrhus Frederic Skinner anderzijds, die onderzoek deed naar dierlijk en menselijk leren door belonen en straffen (in het Engels: *reinforcement learning*) en beroemd, of berucht, werd door zijn theorie over het 'behaviourisme'. Men ging dus programma's maken die 'goede parametersets' belonen en 'slechte parametersets' bestraffen.

Niet alle vroege ideeën trachtten de *mechanismen* achter het menselijk denken na te bootsen. Verschillende wetenschappers concentreerden zich vooral op pogingen om gedragingen of handelingen die door mensen als intelligent beschouwd worden te automatiseren, door gebruik te maken van technieken uit de wiskunde of de statistiek. In 1950 formuleerde Alan Turing een aantal testen waarmee vastgesteld zou kunnen worden of machines kunnen denken. Zijn maatstaf was daarbij het *kunnen beantwoorden van vragen in geschreven taal* op zo'n manier dat de antwoorden van een machine niet meer te onderscheiden zijn van die van een mens. Hij koppelde daarmee het geobserveerde gedrag van de machine los van de manier waarop dit gedrag gerealiseerd wordt: om als intelligent beschouwd te kunnen worden hoeft een machine volgens hem niet te *denken* als een mens, zolang deze zich maar *gedraagt* als een mens. Terwijl de discussies over wat nu precies als intelligentie beschouwd kan worden ook vandaag nog niet afgerond zijn, zette Turing hiermee wel de toon voor het resultaatgerichte perspectief dat vandaag in technische kringen de bovenhand heeft. In deze visie wordt de prestatie van een AI-systeem geëvalueerd op basis van hoe het presteert voor een vooropgestelde taak, ongeacht de manier waarop dit resultaat bereikt wordt. Hiermee wordt bij de ontwikkeling van AI-toepassingen dus vertrokken van een beschrijving van het gedrag dat een AI-systeem moet vertonen (ongeacht of we dit nu 'intelligent' kunnen noemen of niet) en houdt men zich bezig met het zoeken naar technische oplossingen om dit gedrag te realiseren.

In 1956 organiseerde McCarthy een wetenschappelijke bijeenkomst in Dartmouth, om verschillende vroege visies op intelligente systemen samen te brengen. Als overkoepelende term voor al deze benaderingen, en om financiering