Duikboten zwemmen niet

De zoektocht naar intelligente machines

Rudy van Belkom
Stichting Toekomstbeeld der Techniek

DUIKBOTEN ZWEMMEN NIET

De zoektocht naar intelligente machines

Deel I in de reeks ‘De toekomst van artificiële intelligentie (AI)’
Keuzes maken in en voor de toekomst

Rudy van Belkom

Stichting Toekomstbeeld der Techniek (STT)

Een handvat voor de toekomst van AI

Door Maria de Kleijn-Lloyd

Artificiële intelligentie (AI) wordt vaak omschreven als een container-begrip dat het vermogen van computers beschrijft om menselijke oordeelsvorming na te bootsen. De menselijke maat, het menselijke oordeel staat dus centraal. Daarmee bedoelen we vaak een ‘betere versie’ van die menselijke oordeelsvorming. Bijvoorbeeld, een routeplanner is alleen een succes als deze tijdwinst oplevert ten opzichte van het handmatige alternatief. Of een algoritme dat foto’s van verdachte moedervlekken beoordeelt op mogelijk melanoom, zal niet in de huisartsenpraktijk worden gebruikt als de huisarts het met het blote oog beter kan.

De twee bovenstaande voorbeelden zijn typisch voor veel AI-toepassingen van de huidige generatie. Deze hebben een aantal zaken met elkaar gemeen. Ten eerste: het gaat om duidelijk gedefinieerde problemen, zoals ‘wat is de snelste route van A naar B’. Ten tweede: er is een enorme hoeveelheid relevante data beschikbaar, vaak ook in *real time*. Ten derde, daarmee verbonden: het gaat om beslissingen die de mens zeer vaak maakt en niet te veel tijd aan kwijt wil of kan zijn. Routinematig dus. In deze situaties kan een goed algoritme met de juiste data sneller, goedkoper en preciezer (want er kan meer informatie worden meegewogen) resultaat geven.

Er zijn enorm veel toepassingsgebieden die aan de drie criteria hierboven voldoen. Het potentieel en de impact van AI is enorm, zowel breed als diep. AI in zijn huidige vorm is geen sciencefiction, maar de realiteit van vandaag en de belofte van morgen. Dit besef is in alle geledingen van de maatschappij doorgedrongen. Er is een besef van groot potentieel, maar ook een van internationale competitie. AI verandert hoe we leven en werken. Denk aan mogelijkheden om betere medische diagnoses te stellen en behandeling op maat te bieden; economische groei; groter dagelijks gemak. Aan de andere kant is men ook bezorgd. Bij de groei van geautomatiseerd datagebruik luiden alarmbellen rondom privacy, verlies van zelfbeschikking, manipulatie, verlies van banen en polarisatie van de samenleving. Het is duidelijk dat we als samenleving ‘iets’ met AI moeten en er ‘iets’ van moeten vinden om zo tot juiste besluitvorming te komen. Dat begint met een gedeeld besef waar we het eigenlijk over hebben als we ‘AI’ zeggen. Bij Elsevier stelden we ons vorig jaar deze vraag. Uit dat onderzoek is gebleken dat verschillende delen van de samenleving – media, onderwijs, wetenschap en industrie – zeer verschillend praten over AI. Ze gebruiken letterlijk andere woorden. En, bij gebrek aan een gedeeld, realistisch beeld halen boude uitspraken makkelijk het nieuws: ‘*Killer robots* nemen de wereld over’, ‘Nooit meer hoeven werken’.

De dialoog naar een hoger plan brengen, in het belang van de samenleving, is nodig. En het is dringend, want de toekomst staat voor de deur, nee, is er zelfs al. Ik hoop dat deze studie een handvat vormt voor een geïnformeerde dialoog in onze samenleving over hoe we de toekomst van AI willen vormgeven met elkaar, op basis van een realistisch begrip van wat AI is en kan worden. Met andere woorden: de menselijke maat in AI formuleren, zodat AI op de juiste wijze ten dienste staat van de mens. Deze dialoog kan niet aan algoritmen worden overgelaten. Menselijk oordeel: juist als de kaders niet vastliggen, niet alle data beschikbaar is en het geen routinematige beslissing aangaat, is de mens aan zet. Ik doe mee. U ook?

Maria de Kleijn-Lloyd | Senior Vice President,
Elsevier Analytical Services
Voorzitter denktank STT toekomstverkenning AI

Aanpak onderzoek

Artificiële intelligentie (AI) staat hoog op de agenda bij bestuurders en beleidsmakers. Ondanks het feit dat er veel geschreven is over het onderwerp, is er nog veel onduidelijkheid over de wijze waarop de technologie zich gaat ontwikkelen in de toekomst en wat de mogelijke impact gaat zijn op de samenleving. Het is daarom tijd voor een toekomstverkenning naar AI. Een groot verschil met andere technologieën is dat AI steeds meer autonoom wordt, waardoor we steeds meer beslissingen uit handen gaan geven. Dit heeft geleid tot de volgende centrale vraag:

Wat is de impact van AI op besluit-vorming in de toekomst?

Menselijke besluitvorming komt tot stand door middel van verschillende componenten. Naast feitelijke kennis spelen de beeldvorming en ambities van de betrokkenen in het besluitvormingsproces een rol. De invulling van deze componenten is sterk aan verandering onderhevig. Daarbij hebben deelnemers aan besluitvormingsprocessen vaak uiteenlopende ideeën over de werkelijkheid. Wat voor de één een onweerlegbaar feit is, is voor de ander discutabel. In dat opzicht zou je kunnen stellen dat *automated decision making* de nodige objectiviteit kan inbrengen. De vraag blijft echter in hoeverre AI menselijke besluitvormingsprocessen kan overnemen. En in hoeverre we dit willen (en kunnen) toestaan.

De toekomst wordt vaak gezien als iets wat ons overkomt, maar het is eigenlijk iets waar we als mens zelf sturing aan geven. De ambitie van deze toekomstverkenning is om met een multidisciplinaire groep experts te werken aan een formulering van de gewenste toekomst van AI en welke rol Nederland hierin kan spelen. Dit heeft geleid tot het volgende drieluik:

Deel I: Voorspellen

Wanneer het over de toekomst van AI gaat lijken er maar twee smaken te bestaan, de utopie en de dystopie. Vaak starten deze discussies met ethische vragen, terwijl de vraag of de technologie deze scenario's in de toekomst daadwerkelijk kan waarmaken wordt overgeslagen. De focus in deel I ligt daarom op de technologie:

Op welke wijze gaat AI zich in de toekomst ontwikkelen?

Hierbij wordt gebruik gemaakt van forecasting. Aan de hand van literatuurstudies en expertinterviews wordt de meest reële ontwikkelingsrichting van de technologie in kaart gebracht (scope 0-5 jaar). Voor dit deel zijn 40 experts geraadpleegd, van AI experts tot psychologen en bestuurskundigen.

Deel II: Verkennen

De ontwikkelingsrichting van de technologie heeft impact op de mogelijke toekomstbeelden. De meningen van experts lopen enorm uiteen. We kunnen daarom niet uitgaan van slechts één toekomstbeeld. De focus in deel II ligt daarom op de uitwerking van mogelijke toekomst-scenario's:

Welke sociaal-culturele implicaties heeft de ontwikke-lingsrichting van AI op besluitvorming en welke mogelijke toekomstscenario's kunnen aan de hand hiervan uiteengezet worden?

Hierbij wordt gebruik gemaakt van scenarioplanning. Aan de hand van focusgroepen en creatieve sessies worden meerdere toekomstscenario's in kaart gebracht (scope 5-10 jaar).

Deel III: Normatief

AI kan de eerste technologie worden die zijn eigen ontwikkeling gaat bepalen. Het zou in dat geval belangrijker dan ooit tevoren zijn om de gewenste voorwaarden te bepalen voor de ontwikkelingsrichting.

Welke toekomst willen we? De focus in deel III ligt daarom op ethische vraagstukken:

> *Welke ethische vraagstukken spelen een rol bij de impact van AI op besluitvorming in de toekomst?*

Hierbij wordt gebruik gemaakt van *backcasting*. Aan de hand van een Delphi-studie en *round table* gesprekken wordt de gewenste toekomst van AI en de positie van Nederland in kaart gebracht (scope 10-20 jaar). Hierbij wordt tevens gekeken naar de paden die er zijn om daar te komen.

Scope

Toekomstverkenningen van Stichting Toekomstbeeld der Techniek (STT) kennen doorgaans een scope van 20 jaar. Deze scope heeft niet tot doel om concrete uitspraken te doen over een specifiek jaartal, maar het geeft aan dat de verkenningen zich niet beperken tot kortetermijn-ontwikkelingen. Het doel is juist om los te komen van de beperkingen in de huidige tijdsgeest. Hiervoor is het nodig om over grenzen heen te kijken en zo de horizon te verbreden en meer toekomstgericht te denken.

Deze verkenning borduurt voort op de inzichten uit eerdere STT-studies naar data, namelijk *Dealing with the data flood* (2002) en *Data is macht* (2017). Data vormt de grondstof voor hedendaagse AI-technieken. De vraag is of dit in de toekomst zo zal blijven of dat de technologie zich in een andere richting gaat ontwikkelen.

» Technologie heeft vaak geen stekker meer. Die kun je er dus ook niet zomaar uittrekken. «

Denktank

Tijdens de verkenning is uitvoerig gebruik gemaakt van de expertise en ervaring van bestuurders en experts uit het werkveld. De volgende multidisciplinaire denktank is geformuleerd:

Marc Burger	Capgemini	CEO
Patrick van der Duin	STT	Directeur
Bernard ter Haar	Ministerie SZW	DG Sociale zekerheid en Integratie
Frank van Harmelen	VU	Professor Knowledge Representation & Reasoning
Fred Herrebout	T-Mobile	Senior Strategy Manager
Marijn Janssen	TU Delft	Professor ICT & Governance
Maria de Kleijn-Lloyd	Elsevier	SVP Analytical Services, Voorzitter denktank STT toekomstverkenning AI
Leendert van Maanen	UvA	Assistant professor Psychological Methods Department
Marieke van Putten	Ministerie BZK	Senior Innovation Manager
Jelmer de Ronde	SURF	Projectmanager SURFnet
Klamer Schutte	TNO	Lead Scientist Intelligent Imaging
Maarten Stol	BrainCreators	Principal Scientific Adviser

Leeswijzer deel I

Niet iedereen is blij met de komst van artificiële intelligentie (AI). Mensen zijn bang dat banen worden overgenomen, of erger nog, dat ‘de machines’ in opstand komen. Tegenstanders van AI krijgen steeds vaker bijval van gerenommeerde wetenschappers en ondernemers. Zo kan AI volgens wijlen Stephen Hawking het einde van het menselijke ras betekenen. Ook Elon Musk gelooft dat AI momenteel de grootste existentiële bedreiging vormt. Volgens Chris Bishop, directeur van Microsoft Research, zorgen deze ‘Terminator-scenario’s’ echter voor een stagnatie in het onderzoek naar de ontwikkelingen en mogelijkheden van AI. Revolutionaire ontwikkelingen die de mensheid verder kunnen helpen zouden volgens Bishop door angst en wantrouwen teniet gedaan kunnen worden.

Waarom deze publicatie?

AI lijkt een van de meest besproken, maar misschien wel minst begrepen technologieën van dit moment. Het is op sommige vlakken ‘dommer’ dan veel mensen denken, maar op andere vlakken juist veel ‘slimmer’. Een computer als minister-president lijkt nog wat vergezocht, maar AI heeft nu al wel degelijk invloed op onze arbeidsmarkt. Daarbij is het belangrijk om te kijken op welke wijze AI zich verhoudt tot menselijke intelligentie. Want om te kunnen bepalen of computers intelligenter kunnen worden dan wijzelf, moeten we eerst bepalen wat intelligentie precies is, hoe het bij mensen werkt en of we dit kunstmatig kunnen benaderen. Voordat mogelijke sociaal-culturele implicaties kunnen worden onderzocht en ethische vraagstukken kunnen worden beantwoord is het dus van belang om de technologie goed te begrijpen. Dit is de voornaamste aanleiding voor deze eerste publicatie. Daarnaast moeten we goed kijken naar de verschillende krachten die van invloed zijn op de ontwikkelingsrichting van AI, zoals politieke en economische belangen. Het is daarom van belang om ook te kijken naar geopolitieke verhoudingen en de machtspositie van grote tech-organisaties, zoals Google en Facebook.

Voor wie is deze publicatie?

Deze publicatie is voor iedereen die meer wil weten over AI. Het doel is om een volledig beeld te geven van de status van AI, nu en in de toekomst. Het is geen lichte lectuur met hapklare *takeaways*. Deze publicatie gaat de diepte in en probeert AI echt te doorgronden. Wil je goed geïnformeerd kunnen bijdragen aan de discussie omtrent AI? En schroom je de filosofische discussie niet? Dan is deze publicatie geschikt voor jou.

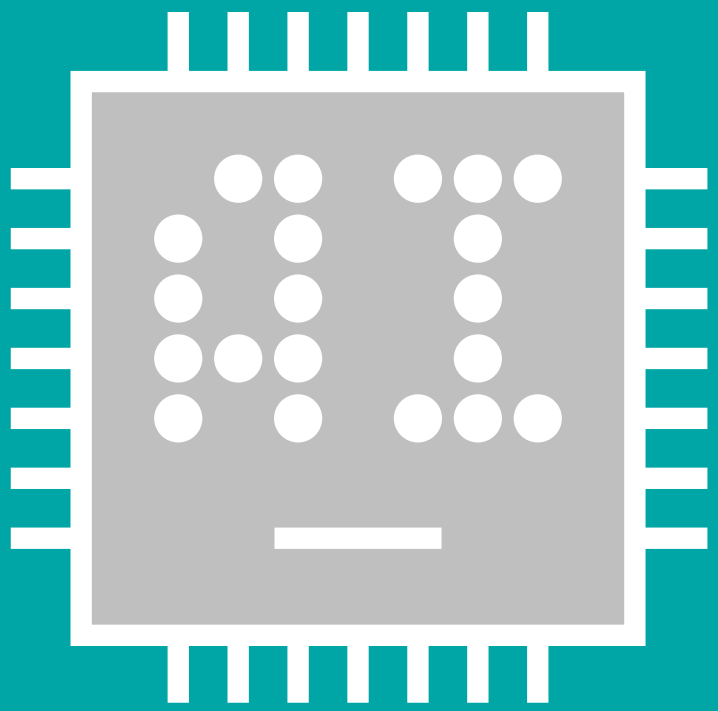
Hoe is deze publicatie opgebouwd?

Deze publicatie kan uiteraard in zijn volledigheid gelezen worden, maar is daarnaast zo opgebouwd dat hoofdstukken en paragrafen afzonderlijk van elkaar te lezen zijn. Het fungeert dus als een soort database, waarbij je zelf kunt bepalen welke delen interessant en relevant voor je zijn. Mocht je toch begrippen tegenkomen die je niet kent, dan kun je eenvoudig terugbladeren.

Hoofdstuk 1: Alles wat je moet weten om te kunnen begrijpen wat AI is en hoe het werkt. Geen complexe wiskundige formules, maar wel een gedegen kijk onder de motorkap en belangrijke lessen uit de praktijk.

Hoofdstuk 2: Alles wat je moet weten om te kunnen bepalen in hoeverre menselijke intelligentie kunstmatig te benaderen is en welke besluiten geautomatiseerd kunnen worden, inclusief de voordelen en beperkingen.

Hoofdstuk 3: Alles wat je moet weten over de ontwikkelingsrichting van AI in de toekomst. Wat vinden de experts? Welke belangen zijn er? En kan AI ons in deze zoektocht helpen?



1 ----------
Wat is artificiële intelligentie? (AI)

Pagina’s: 45
Woorden: 9873
Leestijd: ca. 1,5 uur

Wat is Artificiële Intelligentie (AI)?

Artificiële intelligentie (AI). Het is een begrip waar bijna iedereen wel eens over heeft gehoord. Voor sommigen is het een abstract begrip dat ze kennen van angstaan-jagende krantenkoppen, voor anderen is het dagelijkse praktijk. Wat het niveau van begrip over het onderwerp ook is, de meningen erover lopen enorm uiteen. Niet alleen over de vraag hoe AI zich in de toekomst gaat ontwikkelen en of we menselijke intelligentie gaan benaderen of zelfs overstijgen, maar ook over de basale vraag wat AI nu precies is. Er lijkt geen duidelijke definitie voor AI te zijn die algemeen aangenomen of onomstreden is. Dat maakt de discussie op zijn minst ingewikkeld.

AI is al lang geen sciencefiction meer. Intelligente systemen zijn inmiddels doorgedrongen in ons dagelijks leven, vaak zonder dat we het doorhebben. Denk aan de *spamfilter* in je mailbox, de *recommendations* van je streamingdienst en de *translations* van je zoekmachine. Ook op andere terreinen hebben AI-gedreven tools hun intrede gemaakt. Zo staan vliegtuigen het grootste deel van de vlucht op automatische piloot en zijn *visual recognition* algoritmen beter in het beoordelen van longfoto's dan menselijke artsen. Maar hoe werkt het nu precies?

Hoe werkt AI?

Voordat we de werking van AI kunnen beschrijven is het van belang om het begrip eerst goed te definiëren. AI wordt vaak een *General Purpose Technology* (GPT) genoemd. GPT's zijn technologieën die een volledige economie drastisch kunnen veranderen. Ze hebben de potentie om samenlevingen te ontwrichten door hun impact op bestaande economische en sociale structuren. Denk aan elektriciteit, automatisering en het internet. Toch is AI feitelijk geen technologie. Het is de uitkomst, of het doel, van een ecosysteem van verschillende technologieën. Er kan grofweg onderscheid worden gemaakt tussen drie overkoepelende technologieën die AI mogelijk kunnen maken:

> *Whole brain emulation* (WBE), ook wel *mind upload* genoemd. WBE is het hypothetische proces van het scannen van het volledige menselijk brein, inclusief het langetermijngeheugen, om deze vervolgens te kopiëren naar een computer. Theoretisch gezien zou een computer dan een simulatiemodel van de informatieverwerking van het brein kunnen uitvoeren en in essentie op dezelfde wijze kunnen werken als het originele brein en daarmee bewustzijn kunnen ervaren. *The Human Brain Project* (HBP), een van de twee grootste wetenschappelijke projecten ooit gefinancierd door de Europese Unie, werkt hier al vanaf 2013 aan samen met 500 wetenschappers en meer dan 100 universiteiten, ziekenhuizen en onderzoekscentra.

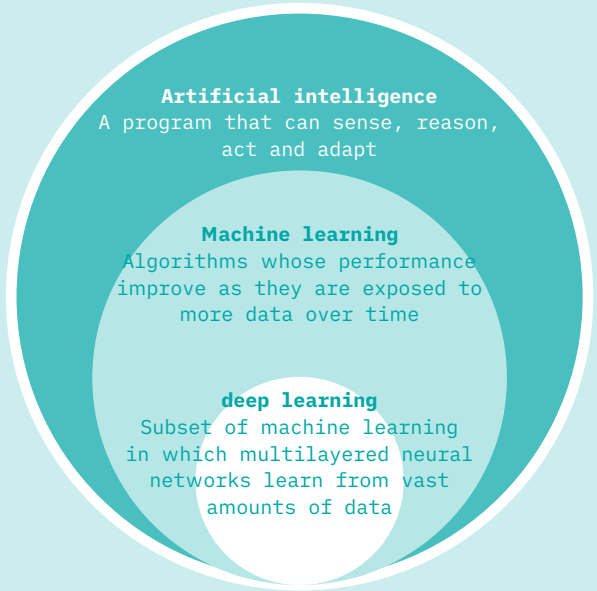
> *Brain-computer interfaces* (BCI) is een technologie waarbij hersensignalen gemeten en gedigitaliseerd kunnen worden. Een computer kan deze vervolgens classificeren en omzetten in acties. Op deze wijze is het bijvoorbeeld mogelijk om robotprotheses aan te sturen.

> *Machine Learning* (ML) gaat over het creëren van algoritmen die kunnen leren van data. Vaak zeggen mensen dat een computer alleen doet waarvoor deze geprogrammeerd is, maar dat is niet altijd meer het

geval. Het gaat bij machine learning over een revolutie waarin mensen niet meer programmeren (als dit, dan dat), maar waarin machines zelf regels afleiden uit data. Een machine learning-algoritme is in staat om zelfstandig patronen uit data te halen, modellen te bouwen en voorspellingen te geven over uiteenlopende zaken zonder voorgeprogrammeerde regels.

Wanneer mensen het over AI hebben, dan hebben ze het in de meeste gevallen over toepassingen op het gebied van machine learning. Dit is zonder twijfel de meest dominante associatie met AI. Uit onderzoek van de *World Intellectual Property Organisation* (WIPO) uit 2019 blijkt tevens dat machine learning de meest dominante AI-technologie is die in patentaanvragen is opgenomen. Binnen deze toekomstverkenning richten we ons daarom hoofdzakelijk op machine learning. Dit lijkt een beperking, maar dat is het zeker niet. De toepassingen van dit vakgebied zijn bijna oneindig breed en dermate complex dat zelfs de AI-experts het niet met elkaar eens zijn.

Een dominant begrip binnen dit veld is momenteel *Deep Learning* (DL). Deep learning is een machine learning-methode die gebruik maakt van verschillende gelaagde *artificial neural networks*. Deze algoritmen zijn geïnspireerd op de wijze waarop neuronen in ons brein werken. Deep learning-algoritmen kunnen uit de voeten met grote hoeveelheden ongestructureerde data en worden veelvuldig ingezet bij toepassingen op het gebied van onder andere beeldherkenning. Deep learning is de belangrijkste *driver* van de successen van AI in de laatste decennia en grotendeels verantwoordelijk voor het ontstaan van de recente ‘hype’ rondom AI.



‘Cousins of Artificial Intelligence’ – Bron: Towards Data Science (2018)

Dé AI-technologie bestaat dus niet. AI bestaat uit een combinatie van technologieën. Je kunt grofweg stellen dat er minimaal drie componenten nodig zijn, namelijk:

- Data: ontwikkeling is big data
- Algoritmen: ontwikkeling is deep learning
- Rekenkracht: ontwikkeling is Graphics Processing Unit (GPU)

De ontwikkeling van deze componenten hangt sterk met elkaar samen. Door de toegenomen beschikbaarheid van data zijn deep learning-algoritmen (die in de jaren ’60 al bedacht zijn) steeds relevanter geworden. Hiervoor zijn processoren nodig die niet alleen sequentiële, maar ook parallelle berekeningen kunnen maken (GPU’s).

De definitie van AI

Het overkoepelende karakter van AI maakt dat het begrip zich niet eenvoudig laat definiëren. Dit blijkt ook uit het advies van het Europees Economisch en Sociaal Comité (EESC). Rapporteur Catelijne Muller stelt in het rapport uit 2017 dat er geen eenduidig geaccepteerde, vastomlijnde definitie van AI is en dat het een containerbegrip is voor een groot aantal (sub)domeinen. De Nationale AI-cursus, een gratis online cursus voor alle Nederlanders, doet een poging het begrip te definiëren:

"AI zijn intelligente systemen die zelfstandig taken kunnen uitvoeren in complexe omgevingen en eigen prestaties kunnen verbeteren door te leren van ervaringen."

Een belangrijke voorwaarde is dus dat het systeem zelfstandig moet kunnen leren. Het is geen statisch model; het kan zich aanpassen aan de praktijk. Met andere woorden, het moet steeds beter worden naarmate het vaker gebruikt wordt. Toch lijkt niet iedereen zich te kunnen, of willen associëren met het begrip AI.

» AI is wat AI'ers doen. «

-- Leendert van Maanen, UvA

Sommige wetenschappers en ontwikkelaars identificeren zich heel sterk met de term AI, terwijl anderen dat beslist niet willen. In de basis doen ze echter hetzelfde werk. Dit zegt dus veel over de reputatie van het begrip 'AI'. AI is een beetje een ongelukkig begrip. De term wordt enerzijds overal opgeplakt (terwijl het soms 'simpele' statistiek is) en soms wordt het er juist vanaf gehaald (bijvoorbeeld bij navigatiesystemen). Dit wordt ook wel the *AI effect* genoemd. We kennen het eigenschappen toe en rekenen het erop af, nog voordat het zich heeft kunnen bewijzen. Tegelijkertijd nemen we het eigenschappen af als deze bewezen zijn.

» AI is whatever hasn't been done yet. «

-- Larry Tesler, 1970

Jarenlang hebben we gedacht dat er voor schaken een hoogstaande, en vooral menselijke, vorm van intelligentie noodzakelijk is. Het inzicht en intellect dat dit vereist zou niet zomaar door een computer geëvenaard kunnen worden. Totdat in 1997 IBM's schaakcomputer *Deep Blue* de wereldkampioen schaken Garry Kasparov versloeg. Vanaf dit moment vinden we schaken niet intelligent meer. Het is 'gewoon' een formalisatie van mogelijke

zetten en een computer met veel rekenkracht kan dit eenvoudig berekenen. Het is niet onmenselijk om onszelf een speciale rol in het universum toe te kennen. Het degraderen van AI-successen helpt ons om onze unieke positie te behouden. Zodra het mysterie is opgelost noemen we het liever automatisering, in plaats van intelligentie.

Het is daarom interessant om naar de verschillende benaderwijzen van AI te kijken. Stuart Russel en Peter Norvig maken in hun standaardwerk uit 2010, Artificial Intelligence: a modern approach, onderscheid tussen vier verschillende benaderwijzen:

> Thinking Humanly

De automatisering van denkprocessen die we associëren met menselijk denken, zoals leren, probleemoplossing en besluitvorming.

> Acting Humanly

Het creëren van machines die handelingen verrichten waarvoor intelligentie nodig is wanneer ze uitgevoerd worden door mensen.

> Thinking Rationally

Het ontwikkelen van berekeningen en modellen die het mogelijk maken om te redeneren.

> Acting Rationally

Het ontwerpen van artefacten die intelligent gedrag vertonen.

Deze verschillende benaderwijzen onderstrepen de complexiteit van AI. Hierbij lopen de begrippen computers, machines en robots vaak door elkaar heen. Toch gaat het in de meeste gehanteerde definities van AI meer over de *'mind'* dan om de *'body'*. Robotica en automatisering zijn niet per definitie AI. Daarom wordt er in deze toekomstverkenning duidelijk onderscheid gemaakt tussen AI en robotisering en richten we ons dus hoofdzakelijk op de software, tenzij anders vermeld.

AI = whatever hasn't been done yet

AI = whatever hasn't been done yet

AI = whatever hasn't been done yet

AI = whatever hasn't been done yet

-- Larry Tesler, 1970

AI = whatever hasn't been done yet

AI = whatever hasn't been done yet

De indeling van AI

Ondanks het feit dat er geen overeenstemming is omtrent de definitie van AI, wordt de indeling van AI wel breed gedragen. Er kan onderscheid gemaakt worden tussen drie verschillende ontwikkelingsniveaus van AI:

> Artificial Narrow Intelligence (ANI)

ANI is een vorm van AI die zeer goed is in het doen van specifieke taken. Denk hierbij aan schaken, aanbevelingen doen en het geven van kwantificeerbare voorspellingen. Ook de huidige toepassingen op het gebied van beeld- en spraakherkenning zijn *narrow*. Eigenlijk zijn alle toepassingen die we momenteel van AI kennen specialistisch. We zijn namelijk nog niet in staat om AI te ontwikkelen die een taak kan uitvoeren die zijn eigen domein overstijgt. Wanneer je een navigatiesysteem bijvoorbeeld vraagt om je veters te strikken zal deze niet in actie komen.

> Artificial General Intelligence (AGI)

AGI wordt ook wel *human-level* AI genoemd. Deze vorm van intelligentie zou in staat moeten zijn om alle intellectuele taken uit te voeren die een mens ook kan uitvoeren. Dit betekent naast bijvoorbeeld patronen herkennen en problemen oplossen ook het flexibel aanpassen in nieuwe omgevingen. Het is de stip op de horizon. Maar hoe verder we erin duiken, hoe meer we beseffen dat we geen goed beeld hebben van hoe menselijke intelligentie precies werkt. Experts ↴ zijn het dan ook niet met elkaar eens. Volgens sommigen is general AI om de hoek, terwijl het volgens anderen nog bijna 200 jaar gaat duren.

> Artificial Super Intelligence (ASI)

ASI kan bereikt worden wanneer AI het kunnen van het menselijk brein op alle mogelijke domeinen overstijgt. Dus ook op het gebied van wetenschap, creativiteit en sociaal gedrag. Ook al lijkt dit nog erg ver weg – we hebben AGI immers nog niet bereikt – zijn er wetenschappers die stellen dat de stap van AGI naar ASI betrekkelijk klein is. Zo zou een *Seed AI*, een vorm van AGI die zichzelf kan verbeteren door zijn eigen codes te herschrijven, voor een *intelligence explosion* kunnen zorgen. Oxford professor Nick Bostrom beschrijft in het boek *Superintelligence; Paths, Dangers, Strategies* hoe we deze explosie van intelligentie kunnen overleven.

De werking van AI

Het idee van intelligente machines stamt al uit de jaren ’50. In 1950 schreef Alan Turing, sleutelfiguur in het breken van de Enigmacode van de Duitsers in de Tweede Wereldoorlog en grondlegger van de computer, een artikel over Computing Machinery and Intelligence. Een belangrijk inzicht hieruit, wat later ook overgenomen is in AI, is dat ons brein en onze *mind* zich op dezelfde manier tot elkaar verhouden als een computer tot een computerprogramma. Dit impliceert dat ons brein een informatieverwerkingssysteem is en dat denken een vorm van computatie, oftewel een wiskundige berekening is. Zijn uitgangspunt was dat je met een computer álles kunt berekenen en dus ook intelligent gedrag kan vertonen. Hiervoor introduceerde hij in hetzelfde artikel de zogenaamde *Imitation Game*, later bekend als de *Turing Test*. In plaats van de vraag of machines kunnen denken, stelt hij de vraag of machines intelligent gedrag kunnen imiteren. Een computer zou voor deze test slagen wanneer mensen niet weten of ze met een mens of een computer communiceren.

> *» Denken is rekenen. «*
>
> -- Frank van Harmelen, VU

AI is in de basis ‘gewoon’ wiskunde. Weliswaar een enorm geavanceerde vorm van wiskunde, maar het blijft wiskunde. Het is boven alles een middel om een optimalisatiedoel te bereiken. Hierbij draait alles om beslissingen. Bijvoorbeeld bij de zelfrijdende auto: het doel is om veilig van A naar B te komen. Om dit te realiseren moeten er continu beslissingen genomen worden. Zo werkt het in een organisatie in principe ook. Vaak is het

doel winstoptimalisatie. De ene keer is de oplossing deep learning, de andere keer volstaat een ouderwetse regressieanalyse. Wanneer je bijvoorbeeld als ijssalon je verkoop wilt optimaliseren, dan is het handig om je verkoop af te zetten tegen het weer. Je zult dan ongetwijfeld tot de conclusie komen dat het weer en de ijsverkoop samenhangen. Je kunt dan dus ook voorspellen dat er bij mooi weer meer ijs verkocht wordt. Is dit AI? Nee, maar het is wel een algoritme als je dat formaliseert.

Een algoritme is dus een wiskundige formule. Het is een eindige reeks instructies die vanuit een gegeven begintoestand naar een vooraf bepaald doel leidt. Zo heeft Facebook er bijvoorbeeld baat bij dat de berichten van haar adverteerders *geliked* worden door gebruikers. Wanneer je dus vaak voetbalberichten liked, dan zal het algoritme steeds vaker vergelijkbare content op je *timeline* naar boven laten komen. Een neuraal netwerk is ook een wiskundige formule en dus ook een algoritme. In het algoritme van een neuraal netwerk zit een component waardoor het zelf patronen uit data gaat leren.

AI heeft zich in de afgelopen decennia in verschillende richtingen ontwikkeld:

Good Old-Fashioned Artificial Intelligence (GOFAI)

In de eerste fase van AI, van 1957 tot eind jaren ‘90, lag de nadruk voornamelijk op *symbolic representations*. Het uitgangspunt hierbij is dat ons brein een symbool-manipulator is en stapsgewijze processen volgt om voorstellingen van de wereld met berekeningen te ontleden. Deze benadering wordt daarom ook wel *Symbolic AI* genoemd. Er wordt hierbij gebruikt gemaakt van *formal logic*, oftewel als-dan regels. Dit betekent dat een computer in staat is om geldige afleidingen te maken uit kennis die de computer al bezit. Dus bijvoorbeeld als de computer weet dat A gelijk is aan B en ook dat B gelijk is aan C, dan moet de computer kunnen deduceren dat A gelijk is aan C. Het werd daarom ook wel een kennisgebaseerd systeem of expertsysteem genoemd, waarbij het systeem specifieke kennis van menselijke experts gebruikt om een specifiek probleem op te lossen. Deze vorm van AI was vooral gericht op *high level cognition*, zoals redenering en probleemoplossing. Dergelijke taken zijn voor mensen zeer moeilijk en men was daarom erg onder de indruk van deze vorm van AI. Schaken is hiervan een perfect voorbeeld. Schaken vindt namelijk plaats binnen een gekaderde wereld, je hebt geen algemene kennis nodig om de regels van schaken te begrijpen en je hoeft de schaak-stukken niet te herkennen. Een abstracte representatie en brute rekenkracht zijn voldoende. Vanuit dat oogpunt was het geen schok dat Garry Kasparov in 1997 werd verslagen door IBM’s Deep Blue.

Machine Learning (ML)

Toen NASA eind jaren ‘90 met een computergestuurd voertuig Mars wilde verkennen, kwamen ze erachter dat deze niet op afstand bestuurd kon worden. De grote afstand zorgde namelijk voor vertraging in de communicatie. Zo ontstond de behoefte aan een autonoom voertuig dat zelfstandig obstakels en ravijnen kon ontwijken. NASA besloot daarom AI-wetenschappers in te schakelen. Zij hadden echter de grootst mogelijke moeite om dit probleem aan de hand van Symbolic AI op te lossen. Wanneer er veel verschillende situaties zijn waar een AI-toepassing voor kan komen te staan, is het bijna onmogelijk om voor al die verschillende situaties de regels in te bouwen in het systeem. Dit probleem wordt ook wel de *explosion of the state action space* genoemd. Er werd daarom vanaf de jaren ’90 naar andere benaderingen gekeken. Een belangrijke wetenschapper in deze ontwikkeling is Rodney Brooks. Hij schreef in 1990 het artikel Elephants Don’t Play Chess en introduceerde hierin de *subsumption architecture*. Hierbij zat de representatie niet in de robots, maar was de omgeving zelf de representatie waarin ze leerden te manoeuvreren. In plaats van het handmatig programmeren van duizenden regels, leerden machines om zelf regels af te leiden uit grote hoeveelheden data.

Hierdoor ontstond een paradigmaverschuiving van *Symbolic*, naar *Subsymbolic AI*. Subsymbolisch betekent dat er niet langer regels zijn die in woorden zijn uit te drukken, maar dat het systeem van veel verschillende voorbeelden geleerd heeft wat de regels zouden moeten zijn. In deze nieuwe fase van AI, ook wel *Nouvelle AI* genoemd, lag de focus in tegenstelling tot de Symbolic AI *op low level cognition*, zoals perceptie en patronen herkennen. Dit vereist grote hoeveelheden data en *training sets*. Binnen machine learning zijn er verschillende manieren voor AI-systemen om te leren. Er wordt onderscheid gemaakt tussen drie typen:

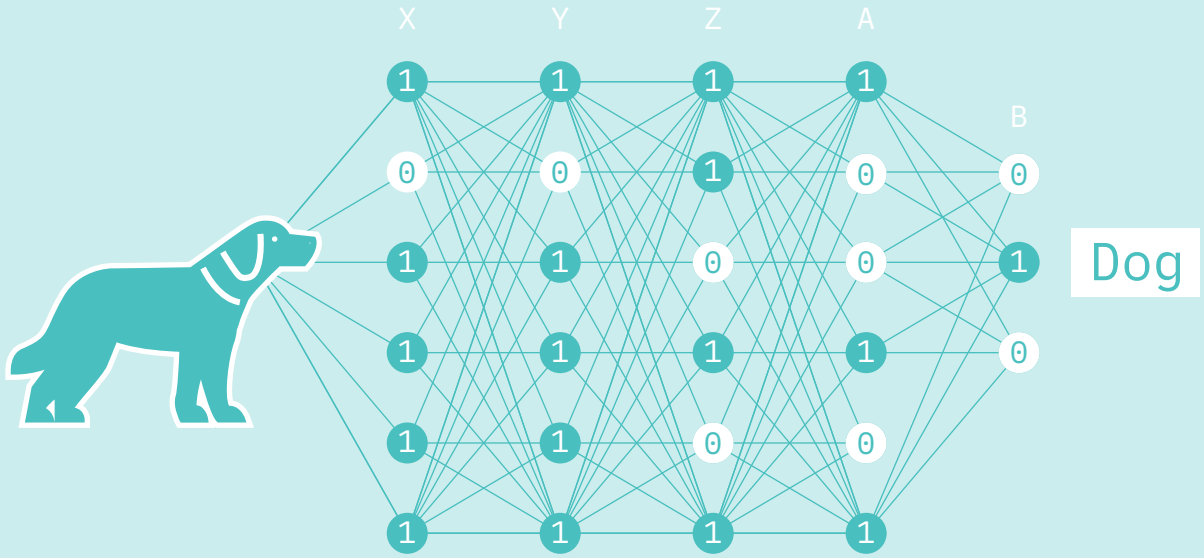
1: Supervised Learning gaat over het zorgvuldig samenstellen en vooraf labelen van trainingsdata. Het algoritme leert dan data te classificeren aan de hand van gelabelde voorbeelden. Wanneer je een algoritme bijvoorbeeld wilt leren om een auto te herkennen, dan kun je deze voeden met duizenden voorbeelden van auto's met het label 'auto'. Gelijktijdig zou je een groot aantal beelden zonder auto's moeten labelen met 'geen auto'. Het helpt om dan verwante beelden te gebruiken, zoals van een tractor en een bus. Wanneer het algoritme getraind is kan deze op basis van nieuwe beelden aangeven of het wel of geen auto is. Hetzelfde principe geldt voor teksten. Bijvoorbeeld spamfilters en vertaalmachines werken op basis van supervised learning.

2: Reinforcement Learning gaat over leren door middel van *trial and error*. De voorbeelden worden hierbij niet vooraf gelabeld, maar het algoritme krijgt op basis van de resultaten achteraf feedback. Wanneer het algoritme de juiste oplossing aandraagt krijgt deze een 'beloning'. Het voordeel van reinforcement learning is dat het systeem processen kan aansturen. Zo wordt het vaak toegepast bij het trainen van spelsystemen, zoals schaakcomputers. Spelsituaties kunnen eenvoudig gesimuleerd worden in een computer en op hoge snelheid afgespeeld worden. Ook de software van zelfrijdende auto's kan op deze wijze getraind worden.

3: Unsupervised Learning gaat over algoritmen die leren om zelf naar clusters te zoeken in grote hoeveelheden ongelabelde data. Het algoritme gaat dan zelfstandig op zoek naar overeenkomsten in de data en probeert op deze wijze patronen te herkennen. Vooraf wordt enkel een maat aangegeven waarmee de afstand tussen *data samples* gemeten kan worden. Het algoritme structureert de data zodanig dat data samples met een kleine afstand bij elkaar in een cluster terecht komen. Dus auto's bij auto's en katten bij katten. Unsupervised learning kan tevens op basis van gebruikersgegevens groepen mensen met vergelijkbare voorkeuren en eigenschappen clusteren. Veel aanbevelingssystemen, zoals de recommendations bij Spotify en Netflix, werken op deze wijze.

Deep Learning (DL)

Binnen machine learning is deep learning een veelbesproken methode. Deep learning is een modelleringmethodiek waarbij neurale netwerken centraal staan. Deze *artifical neural networks* zijn oorspronkelijk gebaseerd op het menselijk brein, waarbij neuronen laagsgewijs met elkaar verbonden zijn. Artificiële neurale netwerken zijn samengesteld uit gestapelde lagen van 'knooppunten' die met elkaar in verbinding staan. Voordat een neuraal netwerk gevoed wordt met data, worden er *random* waardes tussen de 0 en 1 toegekend aan de verschillende knooppunten. Deze waardes representeren de intensiteit van de verbindingen. Wanneer het netwerk gevoed wordt met data wordt de intensiteit van de verbindingen door het netwerk gefinetuned. Deep learning werkt net als neuronen in het menselijk brein die bij specifieke signalen geactiveerd worden en informatie doorsturen. Hierdoor kan het netwerk bijvoorbeeld afbeeldingen van honden classificeren of clusteren.



Visualisatie van een artificieel neuraal netwerk

Er wordt onderscheid gemaakt tussen *input layers*, tussenliggende *hidden layers* en *output layers*. Bij de input wordt de afbeelding opgebroken in pixels. Verschillende layers zijn in staat om verschillende eigenschappen te herkennen. Op een niveau leert het systeem onderscheid te maken tussen allerlei soorten lijnen, strepen en vormen. Op een ander niveau leert het systeem om onderdelen te herkennen, zoals oren en ogen. Hierdoor kan het systeem verschillende beelden van elkaar onderscheiden. Deep learning is een tegenhanger van logisch redeneren en neigt meer naar een vorm van intuïtief leren. Het nadeel hiervan is dat de overwegingen die binnen het systeem gemaakt worden een stuk minder goed te herleiden zijn. Daarbij komt dat een deep learning-netwerk representaties van de data kan maken die voor mensen niet meer herkenbaar zijn.

Bij toepassingen op het gebied van beeldherkenning wordt vaak gebruik gemaakt van een zogenaamd *Convolutional Neural Network* (CNN). Een CNN werkt als een filter die over een afbeelding beweegt en op zoek gaat naar de aanwezigheid van bepaalde eigenschappen. De architectuur hiervoor is weliswaar vooraf ontworpen, maar de filterwaarden worden zelfstandig en automatisch geleerd. Dergelijke neurale netwerken worden ook wel *Feedforward Neural Networks* (FNN) genoemd. De verschillende knooppunten staan weliswaar met elkaar verbinding, maar ze vormen geen cyclus. Ze communiceren dus enkel in voorwaartse richting met elkaar en slaan daarbij niets op. *Recurrent Neural Networks* (RNN) daarentegen zijn in staat om hun interne status, oftewel geheugen, te gebruiken om reeksen van invoer te verwerken. Bijvoorbeeld het *Long Short-Term Memory* (LSTM) netwerk maakt gebruikt van zogenaamde feedback connections waardoor het mogelijk is om een *loop* te creëren. Hierdoor is het in staat om naast enkelvoudige datapunten, zoals afbeeldingen, ook sequenties van gegevens, zoals spraak en video, te kunnen verwerken.

Een doorbraak binnen deep learning waarbij minder data nodig is, is het *Generative Adversarial Network* (GAN). Een GAN kan gezien worden als twee deep learning-netwerken die in competitie zijn met elkaar. Het is een vorm van *semi-supervised learning*. Een van de twee netwerken wordt de *generator* genoemd. Deze probeert nieuwe datapunten te genereren die de trainingsdata nabootsen. Het andere netwerk, de *discriminator*, haalt deze nieuwe data binnen en bekijkt of deze onderdeel zijn van de trainingsdata of dat het *fakes* zijn. Hierdoor ontstaat een positieve *feedback loop*, aangezien de discriminator beter wordt in het onderscheiden van originele en fake data en de generator beter wordt in het creëren van overtuigende fakes.

Binnen deep learning zien we steeds meer combinaties van leren ontstaan. Bijvoorbeeld *Deep Reinforcement Learning*. Hierbij worden bijvoorbeeld drones getraind om op een pad te blijven. Het is een combinatie van *visual navigation en reinforcement*. Zelfs op het gebied van ouderwetse spelletjes levert dit indrukwekkende doorbraken op. Google's DeepMind creëerde *AlphaZero*, een programma die zichzelf *Chess*, *Sogi* en *Go* leerde spelen enkel door tegen zichzelf te spelen. Het algoritme werd niet gevoed met duizenden voorbeeldpotjes, maar werd alleen de regels

uitgelegd. Binnen een trainingsperiode van een dag wist deze de beste computerprogramma's te verslaan.

Het principe van neurale netwerken is overigens allesbehalve nieuw. Al in de jaren '60 werd het principe voor het eerst beschreven. Door de beperkte beschikbare rekenkracht bleven successen echter uit en heeft het lange tijd in de ijskast gestaan.

Functionele toepassingen van AI

AI is in feite een *toolbox* met verschillende soorten gereedschap. Deze tools kunnen worden toegepast op verschillende gebieden, zoals robotica, spraakherkenning en aanbevelingssystemen. In sommige gevallen worden toepassingsgebieden als op zichzelf staande technologieën beschreven, aangezien er overlap is tussen onder andere computertechnologie, cognitieve wetenschap en psychologie. Voor deze toekomstverkenning is er echter voor gekozen om deze toepassingsgebieden als onderdeel van machine learning te benoemen, aangezien machine learning-methoden de belangrijkste drivers zijn van deze technieken. AI kan daarbij per definitie niet zonder inzichten uit andere aangrenzende vakgebieden en is domeinoverstijgend. Voor het overzicht is een selectie gemaakt van de meest besproken toepassingsgebieden van AI:

> *Natural Language Processing* (NLP) gaat over de vaardigheid van een systeem om menselijke taal te begrijpen. Een belangrijk aspect hierbij is semantiek, oftewel de betekenis van symbolen die de bouwstenen vormen van natuurlijke talen. Toepassingen vind je in *chatbots*, vertaalmachines, spamfilters, automatisch gegenereerde samenvattingen en zoekmachines.

> *Speech Processing* gaat over systemen die spraaksignalen kunnen herkennen en verwerken. Dit maakt stemgestuurde zoekopdrachten mogelijk. Denk hierbij aan de verschillende spraakassistenten zoals *Siri* en *Alexa*. Er kan onderscheid worden gemaakt tussen spraakherkenning en stemherkenning. Met stemherkenning kunnen personen door middel van hun stem geïdentificeerd worden. Het is tevens in staat om kunstmatige stemmen te creëren.

> *Computer Vision* gaat over systemen die in staat zijn om informatie uit beelden te halen en deze zo te herkennen. Het doel is dat deze systemen uiteindelijk beelden kunnen begrijpen en inhoudelijk kunnen interpreteren. Beeldherkenning wordt al in verschillende domeinen toegepast, zoals de medische wetenschap (o.a. screenen van longfoto's op afwijkingen), multimedia (o.a. visuele zoekmachines) en defensie (o.a. surveillance). Daarnaast is het mogelijk om beelden kunstmatig aan te passen en zelfs te creëren. Dit worden ook wel deepfakes ↴ genoemd.

> *Affective Computing* gaat over systemen die emoties kunnen opsporen en herkennen. Met name in de marketing wereld is deze toepassing in opkomst. Veel aankopen zijn namelijk emotiegedreven. Ons gezicht en onze houding verraden vaak onze emoties. Het kan daarmee ook worden toegepast binnen politiewerk. Zo zou een overvaller op straat herkend kunnen worden, nog voordat hij of zij de misdaad begaat.

Do(n’t) believe the hype

AI ↴ wordt steeds vaker als een marketingterm gebruikt. Er worden dan ten onrechte populaire termen geplakt op dingen die geen AI zijn. Hypewoorden zorgen voor een overschatting, maar ook voor een onderschatting van de technologie. Als we de media moeten geloven zal AI in de toekomst al onze banen gaan overnemen en de mens volledig buitenspel zetten. Hierbij wordt nauwelijks gekeken naar hoe reëel deze situaties zijn vanuit een technologisch perspectief. Tegelijkertijd hebben mensen vaak niet door dat AI al in steeds meer toepassingen zit die we dagelijks gebruiken.

» We overschatten general AI ↴ *en we onderschatten* narrow AI ↴*. «*

Veel mensen denken nog steeds dat als je een slim recept bedacht hebt, dat daar dan vanzelf de juiste oplossing uitrolt. De intelligentie zit echter niet alleen in de algoritmen, maar ook in de datasets. Er zit veel moeite in het voorbereiden van de juiste datasets om Machine Learning ↴ op toe te passen. Dit wordt vaak onderschat. Veel bedrijven ‘willen er iets mee’, maar hebben hun datamanagement niet op orde. Er wordt een hele geschiedenis aan IT systemen meegesleept. Destijds was data nog niet zo belangrijk. Het is dan ook niet eenvoudig en het vergt veel discipline om iedereen binnen een organisatie te houden aan de afspraken omtrent het vastleggen van data.

Don’t believe the hype

In 2017 kreeg robot Sophia het burgerschap van Saoedi-Arabië. Het ‘product’ van Hanson Robotics was daarmee de eerste robot ter wereld met een rechtspersoonlijkheid. Met een menselijk voorkomen en dito naam werd Sophia in de media neergezet als ‘intelligente robot’. Zonder meer een indrukwekkend stukje *engineering*, maar of het de ontwikkeling van AI ten goede komt is nog maar de vraag. Het is namelijk op zijn minst twijfelachtig of deze robot onder AI geschaard kan worden. Het heeft er alle schijn van dat de gesprekken *scripted* zijn. De timing klopt vaak niet helemaal en daarbij zijn sommige antwoorden té gevat. Bijvoorbeeld op de vraag hoe we een *dark future* kunnen voorkomen, antwoord Sophia laconiek dat we ‘te veel met Elon Musk praten en te veel Hollywood films kijken’. Maar om het verband tussen een dark future, Elon Musk én Hollywood te kunnen leggen, zonder te weten dat deze vraag gesteld kan worden, is extréém veel data en *understanding* nodig en zo ver lijkt de technologie nog niet te zijn. Het beeld van een zelfbewuste robot met common sense ↴ is onterecht. Daarnaast is het op zijn minst opvallend te noemen dat juist Saoedi-Arabië het allereerste land ter wereld is dat rechten geeft aan een vrouwelijke entiteit. Het lijkt daarom meer op een publiciteitsstunt en kan de geloofwaardigheid van AI aantasten.

Mag het een onsje minder?

Een vergelijkbaar voorbeeld waarbij de werkelijkheid wat werd aangedikt zagen we bij het AI-experiment dat het *Facebook Artificial Intelligence Research lab* (FAIR) in 2017 stillegde. Twee spraakrobots, Alice en Bob, zouden hun eigen taal ontwikkeld hebben die alleen zij begrepen. Ze zouden het Engels aangepast hebben om efficiënter met elkaar te kunnen communiceren. Aangezien de onderzoekers er niets van begrepen legden ze het experiment stil.

» We do not know what these bots are saying. Once you have a bot that has the ability to do something physically, particularly military bots, this could be lethal. «

-- Kevin Warwick, University of Reading

Wat er niet bij werd vermeld is dat dergelijke experimenten wel vaker worden afgesloten en opnieuw worden opgestart. Ze waren niet daadwerkelijk met elkaar aan het praten, ze vonden een set patronen die ze beide herkennen. Dit leidde echter niet tot de gewenste output.

Daarom legde Facebook het experiment stil, niet uit angst dat ze te slim werden. Neurale netwerken ↴ zijn nu eenmaal niet goed te begrijpen. Elk model creëert zijn eigen representatie van de data, waarmee dat model goed kan werken. Als wij daar als mens naar kijken dan zijn die niet altijd te begrijpen. Maar dat is geen reden om de boel af te sluiten.

» Elk neuraal netwerk creëert een soort eigen taal om de data te beschrijven. Dergelijke talen zijn moeilijk om te begrijpen. «

-- Maarten Stol, BrainCreators

Er zijn natuurlijk wel processen waar we geen grip op hebben en terecht kun je vragen stellen over hoeveel controle we hebben over onze technologieën en processen. Maar hetzelfde kun je vragen over de financiële markt. Hoeveel controle hebben we daarover? Hoeveel controle hebben we over politieke processen? De vragen zijn in de basis terecht, maar die vragen lijken nu met name voor AI gesteld te worden.

Onbekend maakt onbemind

De angst voor nieuwe technologieën is van alle tijden. Plato was voor het begin van de christelijke jaartelling fundamenteel tegen de komst van het schrift. Hij waarschuwde voor het gevaar van de vergetelheid. Door onze kennis op te schrijven zouden we niets meer kunnen onthouden. Dezelfde kritiek wordt in de huidige samenleving geuit tegen de komst van de mobiele telefoon. We zouden niet eens meer in staat zijn om telefoonnummers te onthouden. Onze angst voor machines die de controle overnemen wordt sterk gevoed door Hollywood. Al bijna 100 jaar worden er films gemaakt over *superintelligence gone wrong*. Door ontwikkelingen op het gebied van AI wordt technologie steeds meer autonoom, waardoor het steeds lastiger is om de gevolgen ervan in te schatten. We weten niet precies waar we bang voor moeten zijn, dus zijn we overal bang voor.

They took our jobs!

Een van de grootste angsten rondom technologisering is de angst dat banen worden overgenomen. Deze angst is van alle tijden. In de 18e eeuw werd de spinmachine niet met open armen ontvangen. *Spinning Jenny*, zoals de machine ook wel werd genoemd, was in staat om met acht spoelen tegelijk wol te weven. Uit angst om overbodig te raken bestormden honderden arbeiders in 1789 fabrieken in Normandië en sloegen er 700 Jenny's en andere machines stuk. Nu de eerste autonome bezorgrobots hun intrede doen in ziekenhuizen lijkt een vergelijkbaar principe op te treden. De robots worden niet alleen geschopt en geslagen, maar ook gesaboteerd door ze bijvoorbeeld op te sluiten in de kelder en ze van hun geplande route af te laten wijken. Dit wordt ook wel *Robo-sabotage* genoemd. De afkeer tegen robots lijkt 'aangemoedigd' te worden door verschillende onderzoeken. Zo zou de helft van de banen verdwijnen als gevolg van robotisering. Deze cijfers werden in 2013 naar buiten gebracht door onderzoekers van Oxford University.

Vijf jaar later bleek uit onderzoek van de *Organisation for Economic Co-operation and Development* (OECD) dat het percentage een stuk lager zou moeten liggen. 'Slechts' 14% van de banen zou op de tocht staan. De voornaamste reden voor deze verschillen is dat OECD beroepen heeft benaderd als een samenstelling van verschillende taken, in plaats van als een massief geheel. En hoewel robotisering als gevolg heeft dat taken veranderen, betekent dat niet per definitie dat het beroep als geheel verdwijnt. Sterker nog, technologie kan ook juist banen creëren. Technologische vooruitgang zou per saldo meer banen kunnen opleveren dan er verloren gaan. Dit wordt ook wel het *ripple effect* genoemd. Een voorbeeld hiervan zie je in de sector van *renewable energy*. Uit onderzoek van de *International Renewable Energy Angency* (IRENA) blijkt dat in 2017 het aantal mensen dat in deze sector werkt met meer dan 5% steeg, wat betekent dat er wereldwijd nu meer dan 10 miljoen mensen in de renewable energy werken.

If we are not careful,
the world will be educating

[v] 2nd-class robots

[ ] 1st-class humans

-- Andreas Schleicher, OEDCD

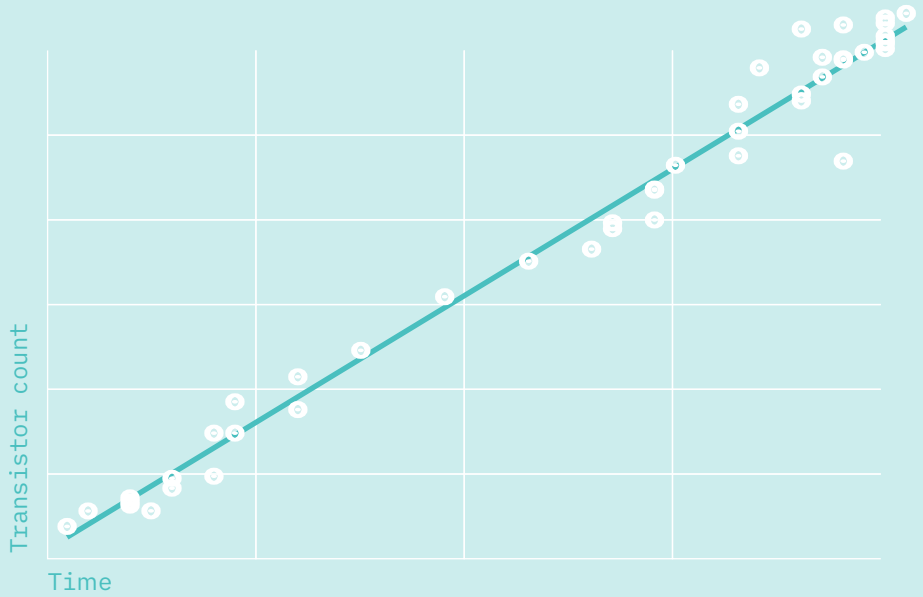
» *We're scared that human jobs will be replaced by robots. But we're still teaching kids to think like machines. If we are not careful, the world will be educating second-class robots and not first-class humans.* «

-- Andreas Schleicher, OEDCD

De inzet van AI binnen bedrijven is momenteel overigens nog erg beperkt. Uit onderzoek van Microsoft uit 2018 blijkt dat 71% van de Nederlandse organisaties AI een belangrijk onderwerp vindt, maar dat slechts 4% van diezelfde organisaties AI daadwerkelijk inzet in verschillende processen en bij meer geavanceerde taken. Bij de meeste bedrijven is AI vooral een voornemen voor de toekomst.

Moore betekent niet altijd 'more'

Moore's Law wordt vaak aangehaald om de exponentiële groei van computertechnologie te duiden. De wet, of eigenlijk voorspelling, stelt dat het aantal transistors in een geïntegreerde schakeling (oftewel chip) door de technologische vooruitgang elke twee jaar verdubbelt.



Visuele weergave van Moore's Law

Deze voorspelling richt zich voornamelijk op *de computing power*, oftewel de rekenkracht en de gegevensopslag. Het zegt in principe niets over de *capabilities*, oftewel de toepassingsmogelijkheden die de toename in kracht oplevert. Een verdubbeling van de rekenkracht leidt niet per definitie tot een verdubbeling van de mogelijkheden. Net als bij geluid; een verhoging van 1DB betekent een 10 keer sterker geluid, maar we ervaren het niet als 10 keer harder. De lijn eenvoudig extrapoleren kan daarbij niet zomaar. Bij grotere complexiteit komen ook grotere uitdagingen. Enerzijds zijn er fysische barrières en anderzijds verandert de industrie en de behoeften daarbinnen. Naast digitale functies worden er ook steeds vaker analoge functies, zoals sensoren en antennes in de chips geïntegreerd.

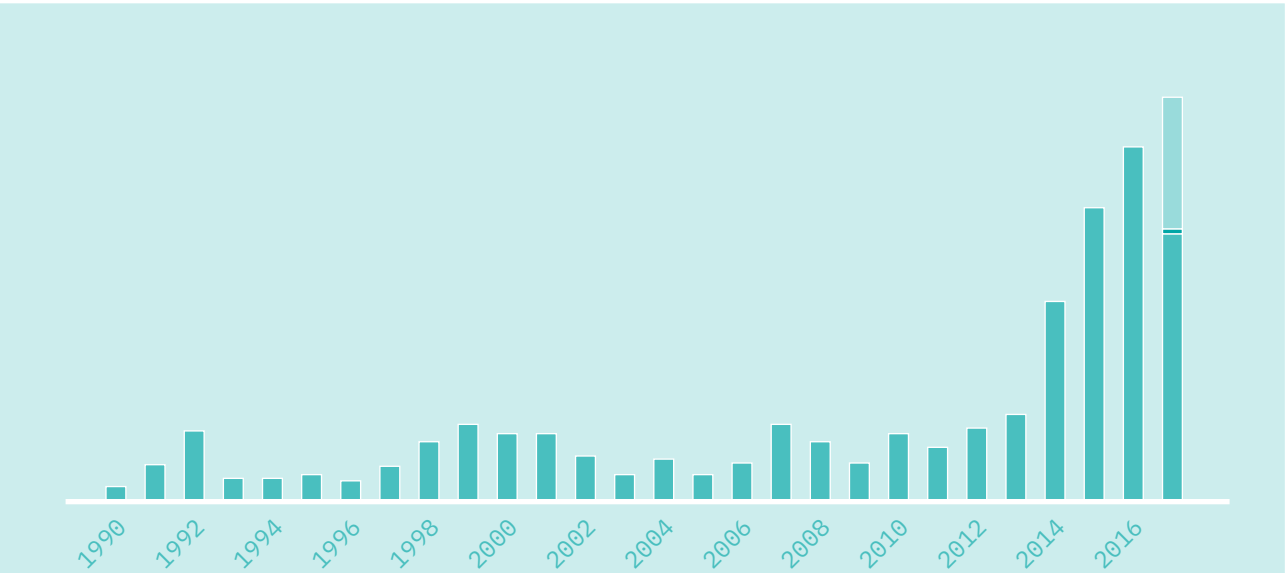
De misbruikte hype

Steeds meer bedrijven willen meeliften op de hype. Uit onderzoek van Marsh & McLennan Companies (MCC) uit 2019 blijkt dat 40% van de Europese bedrijven die bekend staan als *AI startup* in de basis helemaal geen AI-bedrijven zijn. Uit de cijfers blijkt dat startups die zich met AI profileren meer financiering van investeerders aantrekken dan 'reguliere' startups. Ook hier zien we dat de term 'AI' op dingen wordt geplakt die geen AI zijn. Dit lijkt allemaal vrij onschuldig, maar het gevaar is dat bedrijven veel beloven, maar het uiteindelijk niet waarmaken. Dit zorgt voor teleurstellingen, waardoor innovatie in AI geremd kan worden.

Een derde AI-winter?

Wanneer een technologie *overpromises but underdelivers* is dat niet goed voor de ontwikkeling ervan. Het moge duidelijk zijn dat AI allesbehalve nieuw is. De term *'Artificial Intelligence'* werd voor het eerst gebruikt in het proposal for the Dartmouth Summer Research Project door John McCarthy in 1955. Een selecte groep wetenschappers kwam tijdens de zomer samen om de eerste stappen te maken in *human-level machines*. Dit wordt daarom gezien als *the birth of AI*.

De ontwikkeling van AI heeft sindsdien al twee zogenaamde *AI-winters* gekend. Dit zijn perioden waarin de interesse in en financiering voor AI-onderzoek laag is. In de jaren ‘70 lag de ontwikkeling rondom AI bijna compleet stil door felle kritiek en gebrek aan geld. Na een opleving door de komst van expertsystemen ↴ kwam de populariteit van de technologie in de tweede helft van de jaren ‘80 opnieuw in de problemen. Na het dieptepunt in 1990 is de interesse in AI geleidelijk toegenomen. Vanaf 2012 heeft de populariteit van AI, en dan met name van machine learning en daarbinnen deep learning ↴, een vlucht genomen. Dit is goed zichtbaar in het aantal publicaties van het NRC Handelsblad met AI als trefwoord.



AI in het NRC Handelsblad – Bron: Arguments for good artificial intelligence (Verheij, 2018)

De vraag is nu hoe reëel het is dat er een derde AI-winter voor de deur staat. Vaak worden de sociaal-culturele implicaties van de technologie over het hoofd gezien. Bijvoorbeeld bij de zelfrijdende auto; de grootste uitdaging zit waarschijnlijk niet in de technologie, maar in de acceptatie ervan. Ethische vraagstukken omtrent deze toepassing zijn enorm complex

en moeilijk op te lossen. Hierbij komen ook allerlei economische belangen kijken. Wanneer alle voertuigen op de weg autonoom rijden, transformeert het zich steeds meer naar een vorm van openbaar vervoer. Het bezit van een auto zou dan minder noodzakelijk zijn, wat een grote impact kan hebben op de status ervan en dus op het prestige van bepaalde automerken. Het kan dus nog wel eens een hele tijd duren voordat volledig autonome voertuigen het straatbeeld domineren. Technologie verandert vaak ook meer geleidelijk. Een ‘trage vooruitgang’ kan voor investeerders een grote teleurstelling vormen, wat een nieuwe AI-winter zou kunnen aanwakkeren.

Do believe the hype

Ontwikkelingen op het gebied van big data, computing power en opslagcapaciteiten hebben AI een nieuwe *kickstart* gegeven. Daarnaast zien we een ontwikkeling waarbij de acceptatie van nieuwe technologieën steeds sneller gaat. AI is steeds meer geaccepteerd in ons dagelijks leven; denk aan chatbots, aanbevelingssystemen en automatisch inparkeren. Technologie voelt steeds meer natuurlijk. Narrow AI wordt vaak ten onrechte *weak AI* genoemd (en general AI op zijn beurt ten onrechte *strong AI*). Op specifieke domeinen is narrow AI ons namelijk al te slim af. Het haalt nu al patronen uit data die mensen nooit zouden kunnen vinden. De kracht en toepassingsmogelijkheden hiervan worden onderschat.

» Het feit dat een vis geen boom kan beklimmen, zegt niets over zijn zwemkunsten. Integendeel. «

De technical fix

Technologie heeft ons in een houdgreep. Wanneer mensen voor problemen komen te staan is het onze eerste natuur geworden om deze problemen op te lossen met de ontwikkeling van nieuwe technologie. Wanneer we bijvoorbeeld denken over een oplossing om de verkeersveiligheid op de weg te verbeteren, dan zullen we allereerst investeren

Het feit dat een vis
geen boom kan beklimmen,
zegt niets over zijn
zwemkunsten.
Integendeel.

in de ontwikkeling van veiligere voertuigen (met bijvoorbeeld sensoren die de afstand tussen weggebruikers kunnen meten). Het komt minder snel in ons op om bijvoorbeeld de voorwaarden voor een rijbewijs aan te scherpen. Dit wordt ook wel de *technical fix* genoemd. Daarbij zorgt de ontwikkeling van technologieën weer voor nieuwe problemen die vervolgens om de ontwikkeling van nieuwe technologieën vragen. Het lokt elkaar uit. Auto's komen steeds meer vol te zitten met technologische snufjes, waardoor deze sneller kapot gaan. Hiervoor is weer nieuwe technologie nodig om deze te kunnen repareren. Het daagt tevens criminelen uit om op andere, meer technologische wijze, auto's te stelen. Hiervoor heeft de politie weer nieuwe opsporingstechnologieën nodig. Et cetera.

AI is overal

AI wordt al in tal van sectoren en in uiteenlopende toepassingsmogelijkheden geïntegreerd. Zo zijn er hoorapparaten met algoritmen die in staat zijn om omgevingsgeluiden weg te filteren en slimme beeldherkenningssystemen die de inspectie van het spoor vijf keer efficiënter maken. Zonder dat we erbij stilstaan is AI overal. Naast de bekende toepassingen van onder andere Netflix en Google zijn er ook minder voor de hand liggende sectoren waar de potentie van AI ten volle benut wordt. Neem bijvoorbeeld de politie. In 2017 was er nog veel ophef over de falende plaatsbepalingsystemen (PBS), waardoor agenten onderweg in voertuigen hun locatie niet goed konden bepalen. Inmiddels timmeren ze hard aan de AI-weg en openden ze in 2019 het *'Politielab Artificiële Intelligentie'*. Zo worden machine learning-technieken ingezet om de transportlijnen van drugs te volgen, waardoor patronen tussen chauffeurs, vervoersbedrijven en drugsvondsten sneller herkend kunnen worden. Ook kan AI beter inzichtelijk maken op welke wijze online fraude wordt gepleegd en wanneer en op welke plekken er nieuwe methodieken ontstaan. Denk bijvoorbeeld aan de shift van *phishing* naar whatsapp-fraude. Wetenschappers van nu werken aan de politiemacht van morgen.

Serious business

Er gaat serieus geld om in AI. In de afgelopen 10 jaar is er ongeveer 9 miljard euro geïnvesteerd in AI door Europese bedrijven. De meeste investeringen werden gedaan in de financiële sector, gevolgd door media/entertainment en telecom. Uit onderzoek van StartupDelta, de nationale belangenorganisatie voor startups, blijkt dat de meeste startups op het gebied van AI in 2018 te vinden zijn in het Verenigd Koninkrijk, Frankrijk en Duitsland. Zij zijn verantwoordelijk voor 87 procent van alle investeringen in AI-bedrijven. Waar in Nederland zo'n 102 miljoen euro in AI-startups werd geïnvesteerd, was dat in het Verenigd Koninkrijk maar liefst 1,2 miljard euro. De Nederlandse investeringen zijn volgens StartupDelta veel te laag. Zij zijn bang dat Nederland 'de boot gaat missen'.

> De verwachting is dat de investeringen in AI wereldwijd gaan oplopen tot ongeveer 232 miljard euro in 2025. Dit blijkt uit een onderzoeksrapport van KPMG uit 2018. KPMG benadrukt hierin dat veranderingen van de bedrijfscultuur nodig zijn om de effectiviteit van de investeringen te waarborgen. Alle AI-inspanningen moeten onderdeel uitmaken van een strategie die aansluit bij de overkoepelende strategie van de organisatie. Wanneer bedrijven een meer strategische aanpak hanteren ten aanzien van AI, zou de winst tussen de vijf en tien keer zo groot kunnen worden.

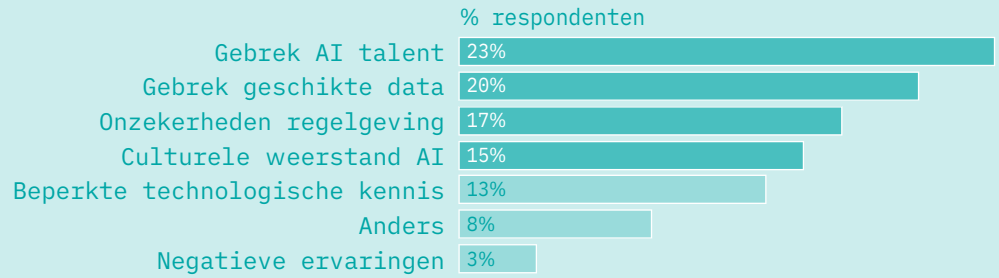
Ook overheden zijn bereid de portemonnee te trekken. Uit de 'Mededeling Kunstmatige Intelligentie voor Europa' uit 2018 blijkt dat de Europese Commissie tot en met 2020 1,5 miljard euro gaat investeren in AI-onderzoek en AI-innovatie. Dit onder andere om de wetenschap en het industriële leiderschap te versterken, maatschappelijke uitdagingen aan te pakken en *AI-research excellence centra* te versterken. Daarnaast wil de Commissie meer particuliere investeringen in AI stimuleren met het Europees Fonds voor Strategische Investeringen. Een enorm bedrag, maar toch maken veel mensen zich zorgen

dat Europa niet kan aanhaken bij Amerika en China, waar veel meer geïnvesteerd wordt in AI. Zo investeert alleen de Amerikaanse Defensie al 1,7 miljard euro in AI-onderzoek.

Omscholen of omvallen

Technologische ontwikkeling gaat historisch gezien gepaard met ontwikkelingen op de arbeidsmarkt. Sommige banen verdwijnen en nieuwe banen ontstaan, maar de meeste banen zullen voornamelijk veranderen van takenpakket. Vaak houden overheden krampachtig vast aan traditionele bedrijfstakken die omvallen, maar in theorie zouden we moeten juichen. Het omvallen van bedrijfstakken biedt namelijk ruimte voor de voor AI relevante bedrijfstakken waar nieuwe banen ontstaan. Als we 1% groei van 'nieuwe banen' willen behalen, moet het aantal werkenden in bestaande banen worden afgebouwd. Dit benadrukt het belang van *upskilling* en *reskilling*, zo blijkt ook uit het rapport 'Arbeid in Transitie' van Denkwerk uit 2019.

Toch kan dit serieuze consequenties hebben die we niet mogen onderschatten. Zeker niet in een tijd van vergrijzing. Een groot deel van de oudere beroepsbevolking is eindelijk gewend aan de digitalisering van veel arbeidstaken. Maar AI gaat een stap verder dan een tabel opmaken in Excel. De vraag is of de complexiteit van AI-gerelateerde taken überhaupt om te scholen is. Uit het rapport 'AI voor Nederland' van AINED uit 2018 blijkt dat een gebrek aan talent nu al als grootste barrière wordt gezien door bedrijven voor het ontwikkelen en gebruiken van AI.



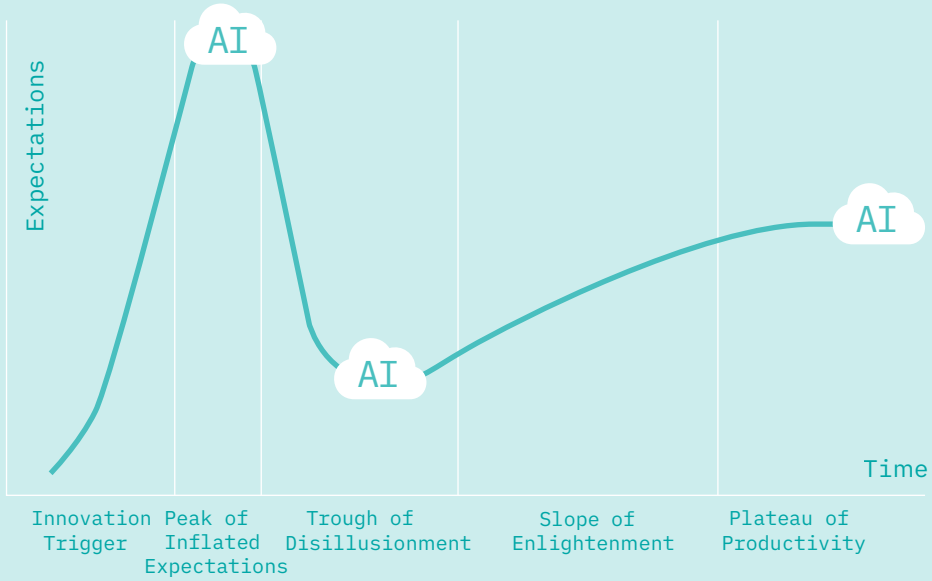
Bron: BCG AI survey voor Nederlandse bedrijven, september 2018

Wanneer het aantal traditionele banen moet worden afgebouwd om de groei in banen te kunnen garanderen, kan er wel eens een enorm gat ontstaan tussen de vraag en het aanbod van *skills* die door het overschot aan oudere arbeidskrachten mogelijkerwijs niet ingevuld kan worden. Ook voor jongeren is het momenteel lastig. Vijf van de zes universiteiten in Nederland die een opleiding in AI aanbieden, beperken de instroom van het aantal studenten of overwegen dat voor het collegejaar 2019-2020 als gevolg van capaciteitsproblemen. De problemen worden voornamelijk veroorzaakt door een gebrek aan docenten. De Cyber Security Raad, een adviesorgaan van het kabinet, vindt dat het Nederlandse Ministerie van Onderwijs extra geld moet uittrekken om een naderende studentenstop te voorkomen.

Voorbij de hype

AI is nu minder een hype dan in de jaren '90. Er zit nu namelijk daadwerkelijk geld achter. Bedrijven als Google investeren enorm, waardoor de technologie daadwerkelijk wordt opgeschaald. Het is dan ook maar de vraag of je AI daadwerkelijk een 'hype' kunt noemen. Technologieën gaan doorgaans de zogenaamde *Gartner Hype Cycle* door. Bij de introductie van een nieuwe technologie stijgen de verwachtingen exponentieel. Het optimisme viert hoogtij.

Maar dan komen de eerste barsten in de euforie. De eerste *failures* zijn een feit. We beginnen ons druk te maken over ethische vragen. Wie is er eigenlijk verantwoordelijk als het mis gaat? De verwachtingen kantelen en geraken in een dal. Het vertrouwen verdampt. Toch gaan de ontwikkelingen (achter de schermen) gewoon door. De grootschalige marktintroductie zal na voldoende tests worden doorgezet en de verwachtingen stabiliseren.
We zullen het gemak boven de mogelijke nadelen verkiezen en accepteren de technologie. Dit klinkt aannemelijk. Toch lijkt deze curve voor AI als overkoepelende 'technologie' niet helemaal van toepassing. AI-technologieën zitten namelijk in alle fases van de cyclus.



Visualisatie Gartner Hype Cycle met positie van AI

AI zit op het hoogtepunt van verwachtingen (de mogelijkheden van deep learning), het zit in het dal van desillusie (de machines nemen onze banen over) en het zit in de gestabiliseerde adoptiefase (uiteenlopende toepassingen in de praktijk). Uiteraard heeft dit te maken met de brede definiëring van het begrip, maar aangezien AI in alle fases van de hype cycle is vertegenwoordigd, lijkt het veilig om te stellen dat het de hype voorbij is.

Lessen uit de praktijk

Hoewel de meeste organisaties nog vooral veel praten over de kansen en bedreigingen van AI ↴, zijn er steeds meer organisaties die AI in de praktijk toepassen. Dergelijke praktijkervaringen zorgen ervoor dat de discussie omtrent AI minder hypothetisch wordt. Het leert ons steeds beter wat AI is, wat het kan en welke uitdagingen er bij implementatie zijn. We hebben daarom twee Nederlandse organisaties gevraagd om de *lessons learned* uit de praktijk te delen. We kijken hierbij vanuit twee rollen:

- De ontwikkelaar: Peter Hulsen, partner bij CQM
- De gebruiker: Kim Larsen, CTIO bij T-Mobile

Beeldherkenning voor spoorinspecties

Door Peter Hulsen, partner

1. Wat was de aanleiding en het doel van het project?

Nederland heeft het drukst bereden spoornet van Europa. Gemiddeld maken reizigers elke dag 1,1 miljoen treinreizen – in totaal 152 miljoen kilometers. Alle goederen leggen met elkaar 6 miljoen kilometer af. Het is dan ook van groot belang zorgvuldig en efficiënt met het spoor om te gaan. Veiligheid en betrouwbaarheid staan hierbij natuurlijk voorop. Door monitoring en inspectie van spoorstaven en -wissels is het mogelijk beginnende defecten eerder te zien en vroegtijdig in te grijpen. AI speelt hierbij een belangrijke rol.

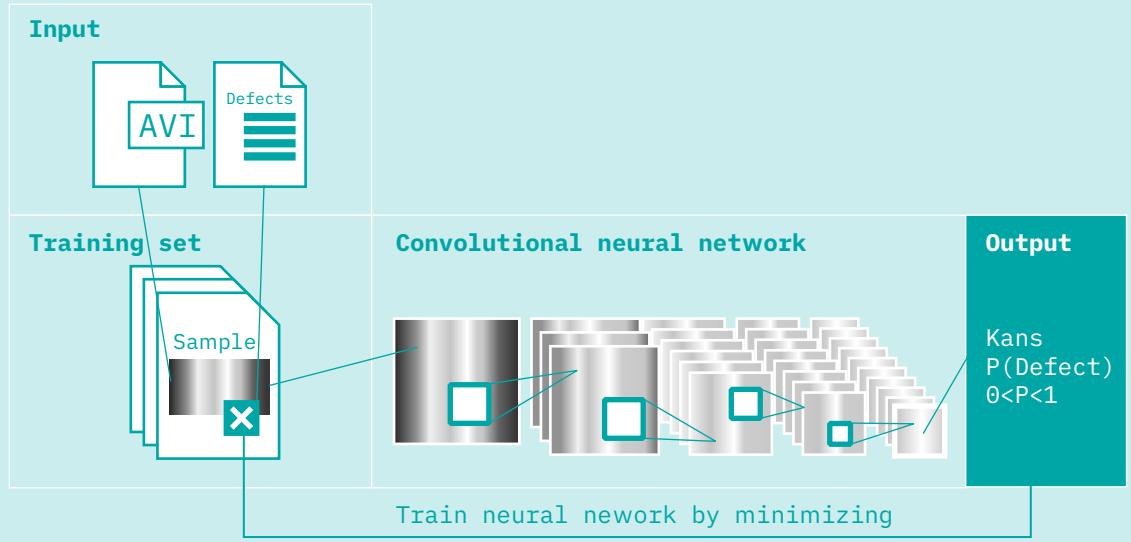
In 2016 zijn Inspectation (nu Asset.Insight, het meet- en inspectiebedrijf van VolkerWessels) en CQM begonnen aan de ontwikkeling van automatische beeldherkenning voor spoorinspecties. Vroeger liepen hiervoor mensen langs het spoor, met alle gevaren van dien. Intussen maken speciaal uitgeruste treinen van ieder stukje rail een foto. Dat gebeurt drie of vier keer per jaar. Dit levert een enorme hoeveelheid beelden op die allemaal door inspecteurs worden bekeken. Met behulp van AI hebben we een zelflerend systeem gemaakt voor het automatisch detecteren van mogelijke defecten op het spoor.

2. Welke AI-technologie is ingezet en waarom is hiervoor gekozen?

Voor de automatische spoorinspectie hebben we deep learning-technieken ingezet. Er is een Convolutional Neural Network (CNN) gebruikt voor het automatisch detecteren van mogelijke defecten op het spoor. Dit type deep learning-netwerk is uitermate geschikt voor het herkennen van patronen in afbeeldingen.

Het CNN is getraind met duizenden foto's die al door de inspecteurs waren beoordeeld. We hebben de foto's opgesplitst in drie sets (training-, validatie- en testsets) om het netwerk te trainen, tunen en testen.

Het netwerk is geïmplementeerd met behulp van het *Google TensorFlow framework*. Berekeningen worden uitgevoerd op speciale hardware, een Nvidia Titan X videokaart. We hebben hierbij een reeds voorgetraind neuraal netwerk gebruikt om de foto's te voorzien van veelvoorkomende kenmerken. En daarmee hebben we een maatwerk netwerk getraind voor het herkennen van specifieke defecten als deuken, krassen, roestvorming en diverse scheuren. Naast het trainen en gebruiken van het neurale netwerk, hebben we een proces gedefinieerd om de kwaliteit te kunnen blijven waarborgen en verder te verbeteren.



Vereenvoudigde weergave van het proces waarbij defecten op het spoor automatisch worden geclassificeerd

3. Wat waren de grootste uitdagingen van de implementatie van de AI-technologie?

Een nadeel van het gebruik van neurale netwerken is dat de beslissingen die genomen worden vaak ondoorzichtig zijn. Het netwerk is een soort black box. Daarom hebben we naast de traditionele *metrics* (zoals accuracy en aantallen (*false positives / negatives*) ook *Gradient-weighted Class Activation Mapping* (Grad-CAM) toegepast om de kwaliteit van het deep learning netwerk te

beoordelen. Hierbij wordt elke foto voorzien van een *heatmap* die de belangrijkste gebieden met betrekking tot het voorspellen van een defect, op de foto markeert. Dit geeft de specialisten een gedetailleerd inzicht in de werking van het neurale netwerk en daarmee vertrouwen in de toepassing ervan. Met name dit soort slimme visualisaties maken het mogelijk om deze complexe, relatief ondoorzichtige technieken toch vol vertrouwen in te zetten in de operationele praktijk.

Het neurale netwerk kon niet direct door een inspecteur worden gebruikt. Uiteindelijk is het getrainde neurale netwerk in de bestaande beeldapplicatie verwerkt. De inspecteurs zien nog steeds dezelfde foto's en inspecteren op de vertrouwde manier. Maar in hetzelfde beeld zien de inspecteurs nu ook een grafiek die door het netwerk wordt aangestuurd. Die grafiek geeft de kans op een defect aan. De inspecteurs hebben nu ook de mogelijkheid om alleen 'verdachte' plekken op het spoor te bekijken waar de kans op een defect boven een bepaalde drempelwaarde uitkomt.

4. Wat heeft de inzet van AI in het project opgeleverd?

Visuele inspectie is een tijdrovende klus en bovendien onderhevig aan interpretaties van verschillende specialisten. Om de kwaliteitscontrole sneller, nauwkeuriger en robuuster uit te voeren kan dit proces met AI deels geautomatiseerd worden. De voordelen zijn drieledig:

- De huidige kwaliteitscontroles kunnen in minder tijd worden uitgevoerd;
- Er ontstaat een objectieve maat voor kwaliteit;
- Bij het opschalen van de productie zal de benodigde tijd voor inspecties niet lineair meestijgen.

Op dit moment wordt de automatische spoorinspectie dagelijks door de inspecteurs gebruikt. Het huidige systeem is zo ver dat de foto's voor 99 procent correct worden beoordeeld. De inspecteurs hoeven 80 procent van de foto's nu niet meer te bekijken. Hierdoor kan een inspecteur vijf keer efficiënter zijn werk doen. De overige foto's bekijken ze nog wel, maar dan om te beoordelen welke acties nodig zijn om de eventuele defecten te repareren.

Naast efficiency levert dit een betere staat van onderhoud op en dus meer veiligheid en beschikbaarheid van het spoor. Doordat de kwaliteit van de inspectie stijgt wordt ook de staat van onderhoud van het spoor beter. We zien beginnende defecten eerder en kunnen daardoor ook eerder ingrijpen. Dit verhoogt niet alleen de veiligheid, maar ook de beschikbaarheid van het spoor. Hierdoor zal er dus ook minder reizigershinder door onderhoud voorkomen.

5. Wat zou je de volgende keer anders doen?

Ondanks dat het project met Inspectation succesvol is verlopen zijn er wel een aantal verbeterpunten. In het vervolg zouden we graag sneller met een grotere dataset het neurale netwerk trainen. Hierdoor is er eerder zorg voor een efficiënte pijplijn voor het bewerken en verwerken van de trainingsdata. Daarnaast geeft een grotere dataset ook direct meer inzicht in de variatie en moeilijkheden in de data. Verder hadden we meer aandacht moeten en kunnen besteden voor de overdracht van de kennis en het neurale netwerk richting de data scientists van Inspectation.

Praktijkcase T-Mobile ----------

AI en telecommunications

Door Kim Larsen, Chief Technology Innovation Officer

1: Wat was de aanleiding en het doel van het project?

De complexiteit van mobiele netwerken wordt steeds groter. Het aantal gebruikers is in de afgelopen decennia exponentieel gestegen en het mobiele gebruikersgedrag is steeds minder voorspelbaar. Daarbij komt dat het netwerk bestaat uit meerdere lagen (2G/3G/4G/5G) en de dichtheid ervan sterk vergroot wordt door wifi. Het handmatig finetunen van de verschillende parameters van dergelijke netwerken is een tijdrovende klus en er kunnen eenvoudig vertragingen optreden bij updates. Deutsche Telekom, het moederbedrijf van T-Mobile, zet daarom steeds meer in op AI. In 2018 is in samenwerking met Uhana, een startup van Stanford University, een cognitieve oplossing geïmplementeerd. De oplossing optimaliseert het zogenaamde *Self Optimizing Network* (SON) zonder menselijke tussenkomst en werkt in real time.

2: Welke AI-technologie is ingezet en waarom is hiervoor gekozen?

Er is gebruik gemaakt van een Recurrent Neural Networks ↴ (RNN), dat getraind is door middel van reinforcement Learning ↴. RNN is een natuurlijke en efficiënte AI-architectuur als het gaat om tijddomein-metingen en -verwerking. Het reinforcement-leren zorgt ervoor dat de optimalisatie van zowel het netwerk, als de klantervaringen gecoördineerd en autonoom gebeurt. Het is de bedoeling om geen menselijke tussenkomst te hebben tussen de Uhana Optimizer, de SON en wijzigingen in de netwerkconfiguratie.

3: Wat waren de grootste uitdagingen van de implementatie van de AI-technologie?

Een van de grootste uitdagingen bij het implementeren van AI-systemen in een zeer complexe live-omgeving zijn mensen. De ervaring met dergelijke oplossingen is beperkt, wat van invloed is op het niveau van vertrouwen in deze oplossingen. Tegelijkertijd zijn er vaak onrealistische verwachtingen als het gaat over complexe AI-toepassingen, die voornamelijk in de beginfase van implementatie moeilijk waar te maken zijn. In dit specifieke geval speelt tevens mee dat de datastromen via de interface van een externe leverancier lopen, wat om een transparante samenwerking en veel afstemming vraagt. Bij het inbedden van autonome systemen in een zeer complexe systeemomgeving moet een organisatie aanvankelijk een relatief hoger operationeel risico accepteren.

4: Wat heeft de inzet van AI in het project opgeleverd?

De principes van autonome netwerkoptimalisatie zijn in de basis geverifieerd. Voor meer complexe optimalisatie-doelen zijn de systemen nog niet volledig autonoom, aangezien het beoordelen van de operationele risico's een doorlopend proces is. Daarbij heeft een beperkte toegang tot de benodigde interfaces van externe partijen voor vertraging van het project gezorgd. Desalniettemin heeft de toepassing van AI in deze context veel potentie. Het kan de kwaliteit van het netwerk vergroten, klantervaringen verbeteren en onderhoudskosten verlagen.

5: Wat zou je de volgende keer anders doen?

Het is essentieel om voorafgaand aan het project uitgebreid in gesprek te gaan met externe leveranciers, met name over de interfaces waarvoor de AI-systemen toegang moeten hebben. De kosten en ontwikkelingstijd voor het openstellen van dergelijke interfaces kunnen zeer aanzienlijk zijn en moeten altijd worden opgenomen in de oorspronkelijke businesscase. Verder kunnen dergelijke projecten niet worden behandeld als zuivere technologieprojecten. Er is een breed scala aan bedrijfsverantwoordelijkheden, risico's en veiligheidskwesties die vooraf zeer zorgvuldig moeten worden besproken. Dit is met name het geval als een bedrijf overweegt autonome intelligente *agent*s in te bedden in hun bestaande complexe systemen.

Conclusies en vooruitblik

AI is een complex begrip. Het is een verzamelnaam voor verschillende technologieën en voor minstens zoveel toepassingsgebieden. De scheidingslijn tussen AI en standaard *analytics* is daarbij niet altijd eenvoudig en eenduidig. Het begrip wordt dan ook te pas en te onpas gebruikt. Tot grote frustratie van AI-wetenschappers en ontwikkelaars. Hoe dichter mensen bij de daadwerkelijke ontwikkeling van de technologie staan, hoe meer de urgentie gevoeld lijkt te worden om de hype te ontkrachten en juist te laten zien wat AI niét kan. Karen Hao, AI reporter bij MIT Technology Review, maakte een handige back-of-the-envelope explainer om te kunnen bepalen of er sprake is van AI of niet.

> AI is overal en tegelijkertijd nergens. Toepassingen op het gebied van narrow AI worden steeds sterker en het is inmiddels al een miljardenindustrie. Tegelijkertijd lijkt het benaderen van general AI nog enorm ver weg en blijven het vooralsnog hypothetische discussies. Hierbij wordt de vraag wat intelligentie nu precies is echter vaak overgeslagen. Want weten we eigenlijk wel hoe intelligentie bij mensen werkt? En in hoeverre is dit artificieel te benaderen? Dit vraagt om een beter begrip van intelligentie en de werking ervan.

Praktijkervaringen leren ons dat de potentie en de voordelen van de inzet van AI groot zijn. AI wordt in steeds meer toepassingsgebieden succesvol ingezet. Tegelijkertijd laat het ook de beperkingen zien. De technologie wordt dermate complex dat we niet zomaar meer onder de motorkap kunnen kijken. En als we dit doen kunnen we steeds minder goed volgen wat er gebeurt. Het is hierbij de vraag in hoeverre we onder onze eigen menselijke ‘motorkap’ kunnen kijken. Kunnen we onze besluitvormingsprocessen automatiseren? En is het antwoord op deze vraag voor elk type besluitvorming gelijk? Het is daarom van belang om te kijken naar de wijze waarop AI zich verhoudt tot menselijke intelligentie en besluitvorming en wat hierbij de voordelen en beperkingen zijn.

De AI-hype: een alomvattende toekomst-belofte

Door Maarten Stol, Principal Scientific Adviser, BrainCreators

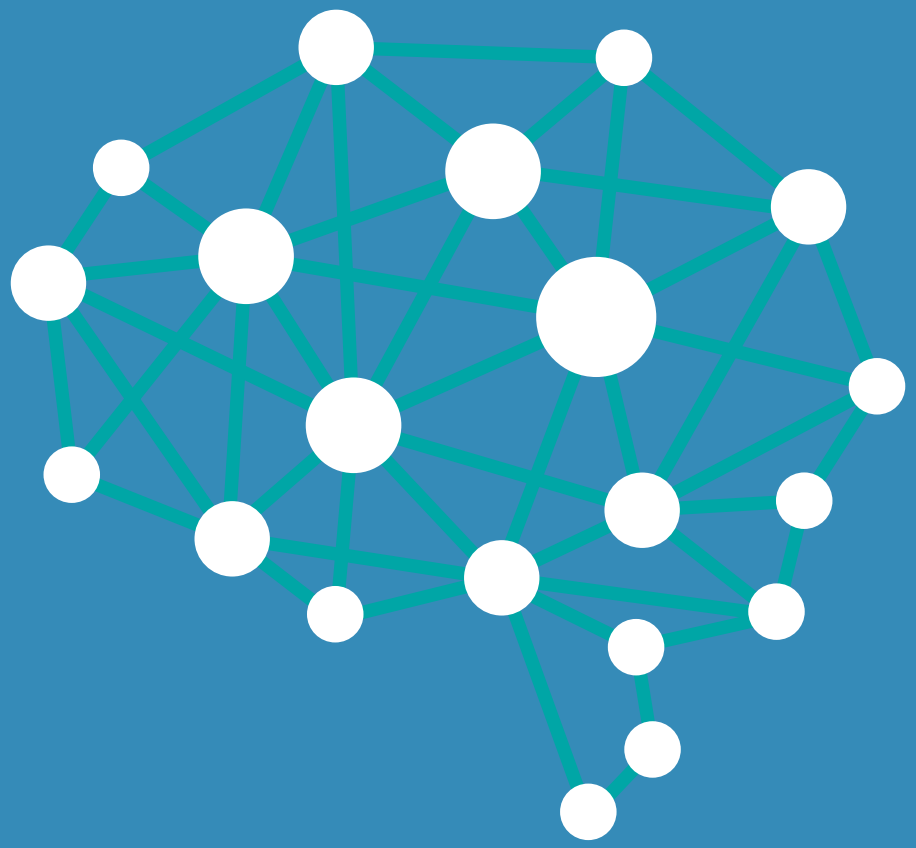
Het is soms makkelijk om te vergeten hoe jong onze moderne wereld is. Elk jaar op vakantie in door robots gemaakte auto's. Betaalbare kleding uit verre landen vandaag besteld, morgen in huis. Internet en social media in je broekzak. Allemaal vanzelfsprekend. En allemaal sinds kort. De snelheid van deze veranderingen geeft een interessante kijk op de huidige hype rond artificiële intelligentie (AI). Illustratief is de veranderende toekomstbelofte waarmee opeenvolgende generaties te maken kregen. Rond 1970 gingen de eerste boodschappen over de voorganger van ons huidige internet, onder andere ontworpen om militaire communicatienetwerken beter bestand te maken tegen een atoomaanval. Beloften voor de toekomst balanceerden dan ook tussen goedkope nucleaire energie en een dreigende nucleaire Apocalyps. Moderne techniek domineerde onze verwachtingen al wel. Informatietechnologie speelde alleen nog geen hoofdrol.

Een generatie later, in 1995, kreeg ik als AI-student aan de Universiteit van Amsterdam mijn eerste e-mailadres. Niet gangbaar onder mijn 'gewone' vrienden. Ook dagelijks toegang tot het internet was nieuw. Klik op een door Tim Berners-Lee bedachte blauwe hyperlink en een document uit Australië stond al op je scherm voordat de thee klaar was. Een omwenteling van de entertainmentindustrie was niet het eerste waar we aan dachten, maar dat dit alles zou veranderen stond vast. Deze toekomstbelofte maakte ons een bevoorrechte generatie: binnen één mensenleven van saai naar sci-fi. Als je het kon bedenken, dan ging je het waarschijnlijk nog meemaken.

Een van mijn vrienden bedacht bijvoorbeeld een website waarop mensen hun hobby's, vakantiefoto's, en karakter-profielen konden plaatsen. AI zou al deze informatie filteren, sorteren, en doorzoekbaar maken. Mensen met gelijke interesses zouden met elkaar in contact gebracht kunnen worden. Bovendien zou alles betaald worden door middel van gerichte advertenties. Ja, mooi idee. Alleen kon dat natuurlijk nooit. We hebben die website daarom niet gebouwd. Tien jaar later hoefde dat ook niet meer. En die omwenteling gaf nog meer lading aan de toekomst-belofte. Voor het eerst kregen we zoiets als technolo-gie-gedreven verwachtings-stress. Een toekomst die alle kanten op kan gaan is een vreemde last om te moeten dragen. Nieuwe markten, nieuwe medicijnen, nieuwe media. Pluk je de vruchten van alle technologie die nog gaat komen, of hangt de tak toch net te hoog?

De hype rond AI is iets vergelijkbaars. Het is de combinatie van een enorme toekomstbelofte voor de generatie van 2020, met een even grote onzekerheid. Sommigen zien AI als de laatste uitvinding van de mens. De uitvinding die alle komende uitvindingen vóór ons gaat doen. Misschien terecht. Al gaat dat nog wel een generatie duren. Interessant is om het ook eens om te draaien. De versnelde technologische ontwikkelingen van de laatste twee generaties hebben een behoefte ontwikkeld aan een alomvattende toekomstbelofte. Dit is de hype rond AI: in de huidige AI proberen we zo'n belofte te herkennen. De laatste uitvinding van de mens moet AI nog worden. De ultieme toekomstbelofte hebben we er al van gemaakt.

AI is een paraplu-begrip dat alle mogelijkheden van digitale innovatie bundelt, en daarmee tevens de verwachtingen, en dus de hype. Dat verklaart ook waarom het zo'n moeilijk begrip is om te definiëren. Met een goede definitie van AI proberen we namelijk niet alleen te beschrijven wat al uitgevonden is, maar tegelijk ook een nog onduidelijke toekomst onder ogen te zien.



2 Hoe verhoudt AI zich tot menselijke intelligentie en besluit-vorming?

Pagina's: 61
Woorden: 14.678
Leestijd: ca. 2 uur

Hoe verhoudt AI zich tot menselijke intelligentie en besluitvorming?

In de eerste fase ↴ van AI ↴ was het uitgangspunt dat ons brein een symboolmanipulator is en stapsgewijze processen volgt om voorstellingen van de wereld met berekeningen te ontleden (als dit, dan dat). Er was in deze fase nog weinig interesse in neurowetenschappen. Met de opkomst van Deep learning ↴ is er in de laatste decennia echter steeds meer aandacht gekomen voor de werking van het menselijk brein bij de ontwikkeling van AI. Bij veel discussies omtrent de vraag of AI menselijke intelligentie kan benaderen of overstijgen wordt echter niet vastgesteld wat de gehanteerde definitie en benaderwijze van intelligentie precies is. Wanneer noemen we iets intelligent? Hoe werkt intelligentie bij mensen? En kunnen we intelligentie automatiseren met AI?

Hetzelfde geldt voor geautomatiseerde besluitvorming. In veel gevallen worden verschillende typen besluitvorming op een hoop gegooid. Terwijl de impact ervan enorm afhankelijk is van de context en het toepassingsgebied. De implicaties van besluitvorming van AI binnen de rechtspraak zijn bijvoorbeeld compleet anders dan bij media- en entertainmenttoepassingen. Het is daarom belangrijk om goed te bepalen hoe menselijke besluitvorming precies werkt en in welke mate dit geautomatiseerd kan worden door AI.

Perspectieven op intelligentie

Intelligentie is een abstract begrip. Twee Nederlandse onderzoekers, Wilma Resing en Pieter Drenth, omschrijven intelligentie als een aaneenschakeling van verstandelijke vermogens, processen en vaardigheden, zoals:

- abstract, logisch en consistent kunnen redeneren
- relaties kunnen ontdekken, leggen en doorzien
- problemen kunnen oplossen
- regels kunnen ontdekken in schijnbaar ongeordend materiaal
- met bestaande kennis nieuwe taken kunnen oplossen
- je flexibel kunnen aanpassen in nieuwe situaties
- zelfstandig kunnen leren, zonder directe en volledige instructie nodig te hebben

Om te kunnen bepalen in hoeverre intelligentie kunstmatig te benaderen is door AI, is het van belang om intelligentie vanuit verschillende perspectieven te bekijken:

> Het neurowetenschappelijk perspectief: intelligentie wordt gezien als het vermogen om iets te kunnen leren en begrijpen. Het is een vorm van cognitie, oftewel het vermogen om informatie op te kunnen nemen en verwerken. Dit vermogen wordt gekoppeld aan ons brein. Het is daarom van belang om de werking van het brein nader te bekijken.

> Het neuropsychologisch perspectief: het is niet eenvoudig om intelligentie af te lezen uit ons brein. Binnen de neuropsychologie worden de functies van ons brein in relatie gebracht met gedrag. De mate van intelligentie wordt vaak gemeten aan de hand van psychodiagnostische tests. Denk bijvoorbeeld aan de IQ-test.

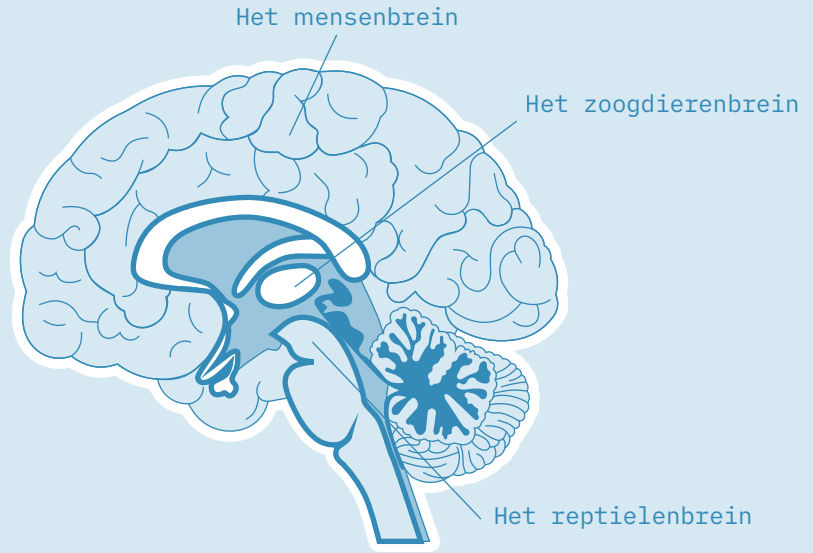
> Het neurofilosofisch perspectief: intelligentie laat zich niet eenvoudig kwantificeren. Er zijn verschillende vormen van intelligentie en het is lastig om deze

exact te definiëren. Intelligentie wordt vaak benaderd als een meetbaar resultaat, terwijl onze vermogens die intelligent gedrag voorstuwen vaak worden overgeslagen. Intelligentie komt onder andere voort uit ons vermogen om onszelf in anderen te verplaatsen.

> Het kunstmatig perspectief: op basis van de inzichten over menselijke intelligentie kunnen we bepalen in hoeverre we dit met AI kunnen bereiken. Computers kunnen goed rekenen, maar in hoeverre kunnen computers daadwerkelijk denken zoals mensen dat kunnen? De vraag is of onze vermogens zoals bewustzijn en empathie kunstmatig te benaderen zijn en in hoeverre dit noodzakelijk is.

Het neurowetenschappelijk perspectief

Het brein is een complex orgaan. Het bestaat uit miljarden neuronen, oftewel zenuwcellen, die met elkaar in verbinding staan. Ondanks het feit dat onze hersenen slechts twee procent van het lichaamsvolume beslaan, verbruiken ze een kwart van de totale energie. Het brein heeft zijn ‘rimpelige uiterlijk’ te danken aan de neocortex (volgens sommigen omdat het eigenlijk niet in de schedel past). De neocortex, Latijn voor ‘nieuwe schors’, is betrokken bij functies die we doorgaans associëren met intelligentie, zoals redeneren, abstract denken en taal. Naast de mens hebben andere zoogdieren, zoals apen en dolfijnen, ook een redelijk ontwikkelde neocortex en vertonen intelligent gedrag. Hierdoor gaat de indeling in ‘mensenbrein’, ‘zoogdierenbrein’ en ‘reptielenbrein’ die vaak wordt gehanteerd niet op.



Visuele weergave van het menselijk brein (zijaanzicht)

Volgens Lisa Feldman Barrett, hoogleraar in psychologie, hangt de mate van intelligentie voornamelijk af van het aantal neuronen en de wijze waarop deze gestructureerd zijn. Ze bepleit hiermee dat de *lizard brain* een mythe is. Mensen onderscheiden zich hoofdzakelijk van andere intelligente zoogdieren door het hoge aantal neuronen (die tevens omgeven zijn door een dikkere isolatielaag, waardoor de signalen sneller doorgegeven kunnen worden).

Ons brein als toekomstvoorspeller

Jeff Hawkins, oprichter van Palm Computing, legt in zijn TED Talk uit 2003 uit dat intelligentie voortkomt uit het vermogen van het brein om voorspellingen te doen. Hij vergelijkt de neocortex met een geheugensysteem. Alle ingekomen informatie wordt opgeslagen in de neocortex. Steeds als er informatie binnenkomt wordt deze vergeleken met hetgeen al is opgeslagen. Wanneer informatie overeenkomt wordt deze niet alleen herkend, er wordt tevens aangegeven wat er bij eerdere ervaringen vervolgens gebeurde. Je hersenen zijn dus eigenlijk continu bezig om de toekomst te voorspellen. Dit gebeurt echter onbewust, het gaat automatisch. Om dit te illustreren introduceert hij het *altered door* gedachte-experiment.

Stel je voor dat iemand de deurknop van je voordeur verplaatst. Wanneer je thuiskomt gaat je hand automatisch naar de plek waar de deurknop hoort te zitten. Het zal misschien een fractie van een seconde duren, maar binnen no time heb je door dat er iets niet klopt.

Je brein doet dit echter niet door een hele database van duizenden deuren af te gaan. Het is intuïtief. Je brein maakt continu voorspellingen over de omgeving. Dit wordt in de cognitieve neurowetenschappen ook wel *predictive coding* genoemd. Hierbij ontstaat een *loop* van voorspellingen over informatie die gaat komen en die wordt vergeleken met de informatie die daadwerkelijk binnenkomt. Bij deze vergelijking kan een voorspellingsfout ontstaan. Deze fout past vervolgens de voorspelling weer aan. Als het verschil niet groot is wordt er weinig aandacht aan besteed, maar als er iets gebeurt wat niet in de lijn der verwachting ligt gaat alle aandacht daar naar uit. Dit verklaart ook waarom je jezelf niet kunt kietelen; de voorspellingsfout is te klein, je weet precies wat je doet.

Het neuropsychologisch perspectief

Neurowetenschappers zijn steeds beter in staat om bepaalde functies aan specifieke hersengebieden te koppelen. Toch weten we over de werking van intelligentie nog relatief weinig. Het is namelijk wezenlijk anders dan bijvoorbeeld motorische functies. Wanneer een specifiek deel in mijn hersenen wordt geprikkeld en mijn arm gaat omhoog, dan is de connectie duidelijk zichtbaar. Bij intelligentie ligt dat anders. De resultaten ervan worden vaak pas later zichtbaar. Het is bijvoorbeeld moeilijk om te bepalen welke zet in een schaakspel precies voor de overwinning heeft gezorgd. Het is meer een samenspel van verschillende slimme zetten. Zo werkt het brein eigenlijk ook; verschillende gebieden werken samen om intelligent gedrag voort te brengen.

Thinking fast and slow

Ondanks het feit dat we nog weinig weten over de lokalisering van hersengebieden die verantwoordelijk zijn voor intelligentie, komen we wel steeds meer te weten over de processen in het brein die verantwoordelijk zijn voor ons denkvermogen. Nobelprijswinnaar Daniel Kahneman introduceert in zijn boek *Thinking, Fast and Slow* uit 2012 het principe van Systeem 1 en Systeem 2. Dit wordt ook wel de *dual processing theory* genoemd.

> *Systeem 1* staat voor het snelle, automatische denken. Het gebeurt onbewust en wordt gekoppeld aan instinctieve en emotionele keuzes. Het is intuïtief en kost ons weinig energie.

> *Systeem 2* staat voor het trage, beredeneerde denken. Het is een bewust proces en wordt gekoppeld aan rationele keuzes. Informatie wordt verzameld en afgewogen en de verwerking ervan kost veel energie.

Er bestaat volgens deze theorie een wisselwerking tussen de twee systemen. Signalen komen binnen bij Systeem 1 die ze vervolgens voorlegt aan Systeem 2. In de meeste gevallen worden de suggesties van Systeem 1 moeiteloos door Systeem 2 overgenomen en vindt de verwerking zonder veel aanpassingen plaats. De meeste informatie wordt dus automatisch en onbewust verwerkt, ook al vinden we dat als rationele wezens moeilijk om toe te geven. We vertrouwen op ons gevoel.

Meten is weten

Om intelligentie vast te kunnen stellen worden intelligentiemetingen gedaan. Dit gebeurt voornamelijk aan de hand van psychodiagnostische tests. Het intelligentiequotiënt (IQ) is daarbij de meest gangbare maat van intelligentie. Het is echter niet eenvoudig om het niveau van intelligentie op een betrouwbare en valide wijze vast te stellen. Intelligentie bevat uiteenlopende vaardigheden die bijna onmogelijk in een test te vangen zijn. Daarbij is het een momentopname en kan de fysieke

en emotionele staat van een deelnemer aan de test de uitkomst sterk beïnvloeden.

Het neurofilosofisch perspectief

Hoe ingenieus de (neuro)cognitieve theorieën ook zijn, er zijn een aantal belangrijke aspecten die ons gedrag voorstuwen en onze keuzes beïnvloeden, die zich niet gemakkelijk wetenschappelijk prijsgeven. Deze kunnen ook wel '*the big 5*' van intelligentie genoemd worden. Achtereenvolgens behandelen we bewustzijn, vrije wil, empathie, creativiteit en *common sense*.

Bewustzijn

Bewustzijn maakt dat we ons onder andere bewust zijn van onze emoties en gedachten, dat we keuzes en de gevolgen daarvan kunnen ervaren en dat we kunnen reflecteren op ons gedrag. Het is daarmee een zeer belangrijk aspect van menselijke intelligentie. Tegelijkertijd is het misschien wel het grootste mysterie. Het onderwerp staat pas vanaf de jaren '80 als serieus onderzoeksgebied op de wetenschappelijke agenda. Een mogelijke verklaring voor het ontstaan van zelfbewustzijn is dat het is geëvolueerd in de eerste grotere apen die zich hoofdzakelijk in bomen verplaatsten. Zij moesten namelijk rekening houden met hun eigen lichaamsmassa ten opzichte van de draagkracht van boomtakken.

Er wordt onderscheid gemaakt in het primair en secundair bewustzijn. Het primair bewustzijn, ook wel het sensorische bewustzijn genoemd, is het vermogen van mensen en dieren om waarnemingen te integreren met geheugen, om op deze wijze een bewustzijn van het heden te creëren. Het gaat dus over het bewustzijn van de wereld om hen heen in het hier en nu, zonder enig gevoel van verleden en toekomst. Het secondair bewustzijn is de toegankelijkheid van een individu tot zijn geschiedenis en plannen en wordt ook vaak geassocieerd met het bewust zijn van eigen bewustzijn. Secondair bewustzijn omvat functies zoals zelfreflectie, abstract denken en wil. De rijkste vorm is aanwezig bij mensen.

Een algemeen geaccepteerde wetenschappelijke verklaring voor ons bewustzijn is er niet. Vaak wordt er gesproken over de *explanatory gap*, oftewel de verklaringskloof. We kunnen weliswaar verklaren op welke wijze pijn fysiologisch werkt (namelijk het afvuren van C-zenuw-vezels), maar dat helpt ons niet om te begrijpen hoe pijn voelt. Pijn is subjectief en moeilijk om objectief over te brengen. Een boeiend fenomeen daarbij is dat wanneer kinderen zich hebben gestoten, ze in sommige gevallen pas gaan huilen nádat ze daadwerkelijk het bloed gezien hebben. Waarneming is niet voor niets een van de belangrijkste onderdelen van bewustzijn. Toch is het bewustzijn van de waarneming niet altijd nodig om ons aan te zetten tot handelen. Wanneer je bijvoorbeeld op straat loopt en op je telefoon aan het appen bent, dan kun je een lantaarnpaal ontwijken zónder dat je deze bewust geregistreerd hebt. Het lijkt dus met name ons onderbewuste (oftewel Systeem 1) te zijn dat ons in leven houdt.

» De inhoud van het bewustzijn is betwijfelbaar, maar het bestaan ervan is zeker, want zonder bewustzijn kun je dat niet betwijfelen. «

Daarbij is het de vraag of bewustzijn deels aangeleerd is of niet. Een baby van 6-12 maanden ziet zichzelf in de spiegel als speelmaatje. Vanaf 18 maanden herkennen baby's zichzelf en hebben dus dan pas een vorm van zelfbewustzijn. Dit is aangetoond door middel van de 'spiegeltest'. Door een vlekje op het hoofd te plaatsen, wordt gekeken of ze eraan voelen of het eraf proberen te poetsen wanneer ze zichzelf in de spiegel zien.

Vrije wil

Een begrip dat samenhangt met bewustzijn en vaak in relatie wordt gebracht met menselijke intelligentie is vrije wil. De vrije wil wordt doorgaans gedefinieerd als het vermogen van rationeel handelende wezens om

controle uit te oefenen over de eigen daden en beslissingen. Toch is de vrije wil bij mensen niet onomstreden. Met het inzicht dat veel van ons gedrag voortkomt uit Systeem 1 processen, wordt aangenomen dat onze daden en beslissingen veelal onbewust zijn en dus niet volledig gecontroleerd worden door onze vrije wil. Deze zienswijze wordt binnen de filosofie ook wel de deterministische stroming genoemd. Het determinisme stelt dat alle gebeurtenissen onder invloed van de natuurwetten volledig bepaald worden door voorafgaande gebeurtenissen en impliceert dus dat de wil niet vrij is. Deze visie wordt gedeeld door gerenommeerde wetenschappers, zoals Dick Swaab. Zij stellen dat het brein materieel is, dat materie onderhevig is aan natuurwetten en er dus geen vrije wil is.

Deze visie staat pal tegenover het libertarisme dat stelt dat het individu de mogelijkheid heeft om meer dan één mogelijke beslissing te nemen onder bepaalde omstandigheden. Aangezien de maatschappij individuen moreel verantwoordelijk houdt voor hun handelingen is vrije wil een vereiste. Voorstanders van het libertarisme stellen dat het determinisme te kort door de bocht is. Er is namelijk geen eenduidige definitie van vrije wil, waardoor men het niet over hetzelfde heeft. Daarnaast is er op geen enkele manier een experiment gedaan dat vrije wil definitief uitsluit. Dat wij gedrag automatiseren wil nog niet zeggen dat we geen vrije wil hebben. Bijvoorbeeld in het geval van een tennisser: het daadwerkelijk reageren op een service is reactief, maar de volgende zet bepalen is strategie. Net als doelen stellen en daarnaartoe werken.

Empathie

Empathie, oftewel inlevingsvermogen, is de vaardigheid om je in te kunnen leven in de situatie en gevoelens van anderen. Door empathie zijn we in staat om ook de non-verbale communicatie van anderen te lezen en te begrijpen. Niet alleen mensen, maar ook sommige diersoorten vertonen empathie. Zo zijn er apen die elkaar troosten en olifanten die terugkomen op de plek waar een familielid is gestorven.

Er zijn echter wetenschappers die de vraag stellen of empathie wel zo positief is bij het nemen van beslissingen. Yale-professor Paul Bloom bepleit in zijn boek *Against Empathy* uit 2018 dat empathie helemaal niet zo'n goede raadgever is als wij denken. Met name wanneer het gaat om langetermijnbeslissingen. Empathie blijkt namelijk enorm eenkennig te zijn. We voelen meer empathie voor mensen die op ons lijken. Daarbij richt het zich vaak op individuen, of kleine groepen. Dat verklaart waarom veel mensen zich drukker maken om het lijden van iemand die van de fiets is gevallen, dan om het lijden van een hele bevolkingsgroep, of het lijden van de planeet.

> *» You can't put yourself in the shoes of a tree. «*
>
> -- Paul Bloom

Mensen die erg empathisch zijn blijken vaak ook enorm wraakzuchtig te zijn. Wanneer er bij een groep mensen (met name uit de eigen sociale groep) onrecht wordt aangedaan, dan moét ertegen worden opgetreden. Zijn conclusie is dat empathie op de korte termijn heel leefbaar is, maar dat het op de langere termijn averechts kan werken. Wanneer er bijvoorbeeld één kind sterft als gevolg van vaccinatie, dan vertelt empathie je dat de hele operatie stilgelegd moet worden. Maar het stoppen van vaccineren kan betekenen dat er veel meer kinderen zullen sterven op de langere termijn.

Creativiteit

Creativiteit wordt vaak gezien als het wonder van het menselijk brein. Het is niet zozeer een specifiek vermogen, maar meer een aspect van menselijke intelligentie in het algemeen. Het is verankerd in conceptueel denken, perceptie, geheugen en reflectie. Creativiteit is het vermogen om ideeën of artefacten te ontwikkelen die nieuw, verrassend en waardevol zijn. Net als bij bewustzijn en vrije wil zijn wetenschappers er nog niet helemaal uit hoe het precies werkt. Hoogleraar psychologie en filosofie Margaret Boden bepleit in het artikel

Creativity in a Nutshell uit 2009 dat we niet zozeer de vraag moeten stellen óf iets creatief is of niet, maar dat we onszelf moeten afvragen hóe creatief iets is en op welke wijze. Ze maakt onderscheid tussen drie vormen van creativiteit:

1: Extrapolatie: bij deze vorm van creativiteit gaat het om het combineren van bestaande ideeën. Deze nieuwe combinaties kunnen zowel bewust, als onbewust ontstaan.

2: Exploratie: deze vorm van creativiteit heeft te maken met het exploreren van denkstijlen. Bijvoorbeeld een schrijf- of schilderstijl. Creativiteit ontstaat hierbinnen door nieuwe ideeën binnen de bestaande denkstijl te ontwikkelen.

3: Transformatie: bij transformatie worden de beperkingen van denkstijlen doorbroken. Dit wordt ook wel gezien als de diepste vorm van creativiteit. Het ontstaat wanneer ‘het ondenkbare’ mogelijk wordt gemaakt. Dergelijke ideeën kunnen enkel ontstaan wanneer denkstijlen worden aangepast, of zelfs volledig getransformeerd worden.

Creativiteit is meer dan ‘nieuwigheid’. Wanneer een kleuter een random opeenvolging van noten op een piano aanslaat, dan zal dit niet direct als creatief gezien worden. Creativiteit wordt ook begrensd door de geschiedenis: wat telt als creativiteit in een bepaalde periode of plaats, kan in een andere periode als belachelijk, dom of gek worden bestempeld. Een gemeenschap moet ideeën dus accepteren als creatief en creativiteit is daarmee sociaal ingebed in onze samenleving.

Common sense

Common sense bestaat uit alle kennis over de wereld; van fysieke en zichtbare aspecten, tot culturele en dus meer impliciete regels, zoals hoe je met elkaar omgaat. Het is voor ons bijvoorbeeld heel logisch dat als we onze sokken in de wasmachine stoppen, dat deze er de volgende dag nog steeds inzitten. En dat als je een vreemde op straat een klap geeft, dat deze dan boos wordt (en je hoogstwaarschijnlijk een klap teruggeeft).

Mensen leren echter niet alleen door feitenkennis, we leren ook door dingen te doen. Door interactie met onze omgeving. Onze omgeving is daarbij geen constante en in veel gevallen cultuurgebonden. Deze kennis wordt ook wel *tacit knowledge*, of onbewuste kennis genoemd. Een belangrijk onderdeel hierbij is het principe van *naive physics*. Dit is het intuïtieve begrip dat mensen hebben van objecten in de fysieke wereld. Baby’s leren al voor hun eerste levensjaar dat voorwerpen niet ophouden te bestaan als je ze niet meer ziet en dat ze niet door een dichte deur kunnen lopen. Mensen hebben een vergelijkbare intuïtie op het gebied van psychologie. Baby’s van drie maanden oud ‘weten’ al dat mensen hun doelen doorgaans proberen te bereiken met zo min mogelijk inspanning.

Intelligentie lijkt vanuit het neurofilosofisch perspectief voornamelijk een eigenschap die we aan onszelf toeschrijven, om de wereld en onszelf begrijpelijk te maken. Met name de complexe concepten die wij aan het brein willen koppelen, zoals bewustzijn en vrije wil, zijn voornamelijk sociale concepten die wetenschappers (overigens met de grootst mogelijk moeite) geconstrueerd hebben. We zijn tot dusver niet in staat gebleken ze te modelleren en kunnen ze niet een-op-een linken aan gebieden in het brein. Dat geldt wel voor aspecten zoals waarneming en motoriek, die heel goed te volgen zijn in het brein. Maar intelligentie is veel moeilijker objectief te meten. Het gaat bij dergelijke metingen toch vaak om interpretaties. We doen hetzelfde met meer alledaagse dingen, zoals honger en dorst. Een ‘flauw gevoel’ is ook slechts een uiting van hetgeen wij honger noemen en feitelijk een tekort aan stoffen in ons fysieke systeem is. Maar omdat deze complexe processen moeilijk te bevatten én over te dragen zijn, hebben we er woorden aan gegeven. Het is een vereenvoudiging. Mensen zouden raar opkijken wanneer je rood beschrijft als ‘de kleur die

je ziet wanneer licht met een golflengte van 700nm door de retina naar je hersens wordt gestuurd'.

Het kunstmatig perspectief

Al vanaf de introductie van de term AI in de jaren '50 staat human level AI ↴ als prominente stip op de horizon.

> *» For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving. «*
>
> -- John McCarthy, 1955

We hebben de neiging om menselijke eigenschappen toe te kennen aan wezens en objecten die niet-menselijk zijn. Dit wordt ook wel antropomorfisme genoemd. Wanneer onze laptop niet opstart, dan zeggen we dat 'hij niet wil opstarten'. Meesters in het spelletje Go noemden Alpha Go 'creatief', nadat ze door het algoritme verslagen waren. Blijkbaar had de computer een bepaald patroon geleerd die de spelers niet kenden. Het lijkt er dus op dat zodra we iets niet meer begrijpen, we er menselijke termen op plakken (net zoals we dat bij complexe eigenschappen van onszelf doen). De verwachting is dat dit alleen maar zal toenemen. Nu is het nog betrekkelijk eenvoudig om de werking en output van machines te verklaren, maar wanneer ze steeds complexer worden ontkomen we er niet aan om over ze te praten in 'geestelijke termen'. Het helpt ons om de wereld overzichtelijk te houden en te begrijpen. Om te kunnen bepalen of computers menselijke intelligentie kunnen benaderen is het van belang om te bepalen in hoeverre *'the big 5'* van menselijke intelligentie geautomatiseerd kunnen worden.

Bewustzijn

Binnen de experimentele onderzoeken over bewustzijn ondervinden wetenschappers met name moeilijkheden vanwege het ontbreken van een algemeen aanvaarde operationele definitie. Het is daardoor vrijwel onmogelijk om een bewijs te leveren die een echt bewustzijn van een kunstmatige vorm van bewustzijn kan onderscheiden. Dit maakt het lastig om zelfbewustzijn van machines volledig uit te sluiten.

Het is daarbij de vraag in hoeverre een hoger bewustzijn voor machines noodzakelijk is om intelligent gedrag te kunnen vertonen. Het gaat met name om de reflectie van de werking van het eigen cognitieve vermogen. Naast *self-awareness* kan dus ook gekeken worden naar *self-assessment* en *self-management*. Een machine kan in dat opzicht een vorm van bewustzijn vertonen wanneer deze in staat is om doelgedreven keuzes te evalueren en deze op basis van ervaringen zelfstandig bij te stellen. Dit komt behoorlijk in de buurt van de definitie van AI ↴.

Het lijkt erop dat we bewustzijn voornamelijk toeschrijven aan andere mensen, omdat hun uiterlijke kenmerken en gedragingen op die van onszelf lijken. Aangezien ik ervan uitga dat ik zelf een bewustzijn heb, ga ik ervan uit dat anderen dat ook hebben. Dit maakt het niet ondenkbaar dat we in de toekomst ook bewustzijn aan machines toeschrijven die op ons lijken.

Vrije wil

Vooralsnog wordt vrije wil hoofdzakelijk toegekend aan mensen. Dit hangt nauw samen met de structuur van onze hersenen. Doordat de meeste dieren niet beschikken over de relatief grote neocortex van de mens zijn zij niet in staat om te reflecteren op hun gedachten en zijn daardoor niet in staat om hun eigen daden en beslissingen te controleren. Zij volgen enkel op wat de natuur heeft voorgeprogrammeerd. Net zoals computers handelen op basis van de optimalisatiedoelen die zijn meegegeven door programmeurs.

Toch kun je je afvragen in hoeverre onze drijfveren daadwerkelijk verschillen van de doelen die we machines meegeven. Emoties, zoals angst, honger, en liefde, zijn ook evolutionair geprogrammeerd om ons te helpen overleven en voort te planten. Het zijn ‘handige trucjes’ van de natuur, zodat wij in termen van overleving de juiste beslissingen maken. Wanneer je een computer programmeert om niet kapot te gaan, kun je dat in feite vergelijken met een mens die niet dood wil gaan.

Empathie

Door empathie zouden mensen in staat zijn om betere keuzes te kunnen maken dan emotieloze machines. Een machine kan niet lijden en kan zich dus ook niet in ons verplaatsen. Maar dat betekent niet dat ze emoties niet kunnen herkennen. Sterker nog, volgens onderzoekers van Ohio State University zijn computers er zelfs beter in dan mensen. Emoties kunnen namelijk afgelezen worden van gezichtsuitdrukkingen. Dit gaat in essentie om patroonherkenning, iets waar deep learning-algoritmen enorm goed in zijn. Computers zijn nu al in staat om 21 emoties af te lezen op basis van gezichtsuitdrukkingen op foto’s. De onderzoekers gingen echter nog een stap verder. Hun hypothese is dat emoties de bloedsomloop in het gezicht veranderen. Elke emotie zorgt voor een andere gezichtskleur. Hierdoor zou je in staat moeten zijn om emoties te herkennen, zelfs bij de afwezigheid van gezichtsuitdrukkingen. Aan de hand van machine learning ↴-methodieken zijn de onderzoekers in staat geweest om verschillende emoties succesvol te classificeren aan de hand van patronen in gezichtskleur.

Machines kunnen vervolgens leren wat de verschillende emoties betekenen en op basis van deze emoties hun gedrag aanpassen (overeenkomstig met de optimalisatie-doelen die meegegeven zijn). Ook al kunnen deze machines emoties niet voelen, kunnen ze wel empathisch handelen.

Creativiteit

Veel mensen denken dat échte creativiteit door computers nooit mogelijk zal zijn. Computers hebben geen hoger bewustzijn en zouden daarom niet kunnen waarderen wat ze voortbrengen. Ze zouden geen aanleiding hebben om hun diepste gevoelens te communiceren met anderen. Daarbij wordt een computer altijd nog geprogrammeerd door mensen en zijn het menselijke parameters die opgevolgd worden.

Toch betekent dit niet per definitie dat computers en creativiteit niet samen kunnen gaan. Computers zijn al in staat om kunstwerken te maken, muziekstukken te componeren en grappen te bedenken. En steeds vaker kunnen mensen het onderscheid met het menselijk werk niet meer maken. Daarnaast zijn computers in staat om combinaties te maken die wij niet konden voorzien, misschien wel juist doordat ze niet sociaal zijn ingebed. Wellicht gaat het hier over het nabootsen van creativiteit, maar dat betekent niet per definitie dat het niet creatief is.

Common sense

Veel van de kennis die wij bezitten is voor ons vanzelf-sprekend en we spreken deze dan ook zelden uit. Dit maakt het voor een computer echter extreem lastig om te leren. Je kunt een neuraal netwerk weliswaar voeden met een miljoen artikelen over alles wat we over de wereld weten, maar mensen schrijven de voor ons vanzelfsprekende dingen niet op. Zoals het feit dat een piramide niet eetbaar is. Aangezien dergelijke kennis voor ons enorm voor de hand ligt, is het lastig om in te schatten wanneer je volledig bent. Een bekend experiment in deze context is de ambitie van onderzoeker Douglas Lenat in de jaren ‘80 om alle menselijke kennis die we over de wereld bezitten in een computer te stoppen. Al deze kennis moest geformuleerd worden in logische als-dan regels. Als je buiten door de regen loopt dan wordt je nat. Als je nat bent dan droog je je af met een handdoek. Et cetera. De belofte was dat ze binnen 10 jaar een machine zouden hebben met meer kennis dan de mens, oftewel *super human knowledge*. Toch heeft dit experiment niet geleid tot AI.

Er zijn ontwikkelingen in common sense door het gebruik van kennisgrafen. Dat zijn een soort databases, maar dan met veel meer verbindingen tussen de verschillende onderwerpen. Het is eigenlijk een oude technologie die onder kennisrepresentatie valt. Met kennisgrafen kun je basiskennis aan AI meegeven, die bijvoorbeeld helpt bij het begrijpen van natuurwetten.

Intelligentie opgesloten in een kamer

Zelfs wanneer je alle kennis over de wereld in een computer programmeert, dan blijft de vraag of een computer daadwerkelijk begrip heeft van zijn handelen. Dit begrip wordt ook wel intentionaliteit genoemd. Er bestaan verschillende gedachte-experimenten die de mogelijke aanwezigheid van intentionaliteit bij computers bevragen.

De Chinese Kamer

'De Chinese Kamer' is een gedachte-experiment van filosoof John Searle uit 1980. Hij bedacht het experiment om aan te tonen dat als een computer zich precies zou gedragen als een mens, deze niet per definitie ook denkt als een mens. Searle uitte hiermee kritiek op een ander gedachte-experiment, namelijk de Turing Test, waarin juist de tegenovergestelde stelling wordt aangenomen. Het gedachte-experiment gaat als volgt:

Iemand die geen woord Chinees spreekt wordt opgesloten in een kamer, waar alleen een boek, schrijfgerei en papier aanwezig zijn. Sommige vellen papier zijn leeg, terwijl anderen vellen met Chinees schrift zijn beschreven en geordend zijn in verschillende stapels. Het boek bevat instructies in de moedertaal van de proefpersoon, waarin wordt beschreven op welke wijze gehandeld moet worden wanneer een vel papier via een gleuf de kamer in wordt geschoven. Afhankelijk van het symbool dat op het vel staat, kan een instructie bijvoorbeeld zijn om een teken op het vel te schrijven, deze op de linker stapel te plaatsen en het bovenste vel van de rechterstapel door de gleuf te stoppen.

De kamer kan hierbij gezien worden als een computer, de proefpersoon als een processor, het boek als een programma en de stapels papier als het geheugen. Deze opzet zou kunnen slagen voor de Turing Test, omdat mensen die met het systeem communiceren niet door zouden hebben of ze met een Chinees sprekend persoon aan het communiceren zijn of niet. Searle vraagt zich af of dit een intelligent systeem is. Begrijpt het systeem daadwerkelijk Chinees? Volgens Searle is het antwoord nee, de proefpersoon volgt slechts de regels op en heeft geen idee waar de conversatie over gaat. Net zomin als het boek of de vellen papier dat weten. Computers zijn daarmee volgens Searle symboolverwerkende machines zonder begrip van hun eigen handelen. Daarmee zou general AI ↴ volgens hem onmogelijk zijn.

Mary's Kamer

Intentionaliteit heeft een interessante eigenschap. Naast het algemene begrip dat wij van onze omgeving en ons handelen hebben, heeft deze tevens een subjectief karakter. Denk bijvoorbeeld aan smaken en kleuren. Voor elk individu is de beleving ervan anders. Dit worden ook wel 'qualia' genoemd, oftewel de kwalitatieve eigenschappen van een waarneming. Qualia kunnen in principe niet letterlijk beschreven worden. Wanneer je bijvoorbeeld de kleur rood omschrijft aan iemand die nog nooit rood gezien heeft, dan zal deze nooit volledig begrijpen wat rood is, omdat hij de ervaring ervan mist. Je kunt bijvoorbeeld wel zeggen dat rood de kleur van vuur heeft, maar iemand die nooit rood gezien heeft zal deze analogie niet begrijpen.

Om het bestaan van qualia aan te tonen bedacht filosoof Frank Cameron Jackson het gedachte-experiment *'Mary's Room'*. Mary is een kleurwetenschapper die alles over rood weet. Ze heeft de kleur echter nog nooit gezien, aangezien ze haar hele leven al in een zwart-witte kamer zit en alle overige informatie uit haar omgeving alleen via een zwart-wit beeldscherm ziet. De vraag is of Mary dan enig begrip heeft van de wijze waarop de kleur rood eruitziet.

» *Als Mary alles weet over rood, weet ze dan ook hoe rood eruitziet?* «

Volgens dit gedachte-experiment dus niet. Pas wanneer ze voor het eerst naar buiten mag zal ze daadwerkelijk ervaren wat rood is. Emoties zijn in dat opzicht ook qualia. Je kunt een machine er alles over leren, maar hij zal ze nooit echt ervaren. Het gedachte-experiment ‘de filosofische zombie’ stelt echter dat het zeer moeilijk is om te bewijzen dat qualia daadwerkelijk bestaan. Wanneer je een zombie met een mes steekt zal deze, volgens de algemene opvattingen, niets voelen. Toch kan de zombie zich wel gedragen alsof hij pijn heeft, door bijvoorbeeld ‘au’ te roepen, terug te deinzen of te vertellen dat het pijn doet. Hij kan namelijk alle kenmerken van pijn geleerd hebben en deze toepassen. Ook al gaan we ervanuit dat hij het niet daadwerkelijk voelt, kunnen we dat moeilijk bewijzen.

De wereld is geen kamer

Wat opvalt is dat veel van deze gedachte-experimenten zich afspelen in een afgesloten kamer. Maar hoe reëel is deze vergelijking? Mensen zouden ook geen intelligent gedrag kunnen vertonen als ze hun hele leven in een kamer opgesloten zouden zitten. Zonder ervaringen in de fysieke wereld zou ik ook niet in staat zijn om common sense te ontwikkelen. Dingen hebben geen betekenis van zichzelf. Dingen krijgen betekenis door de rol die ze spelen in ons leven. Onze hersenen werken eigenlijk zelf ook als een Chinese Kamer. Zenuwcellen weten ook niet wat ze doen en waarom ze dit doen. Ze zijn een schakel in het grotere geheel. Net zoals een processor in een computer. Op basis van voorgaande inzichten kunnen we een aantal voorwaarden benoemen voor intelligentie:

1: Intelligentie zit niet alleen in je hoofd *(embodied cognition)*. Intelligentie wordt gevormd door onderdelen van het gehele lichaam.
2: Intelligentie ontstaat door interactie met de omgeving *(symbol grounding)*. Representaties in onze geest krijgen pas betekenis door de koppeling met sensomotorische ervaringen.
3: Intelligentie is sociaal *(theory of mind)*. Intelligentie hangt samen met het besef dat de eigen opvattingen, verlangens en emoties kunnen afwijken van die van een ander.

Om deze voorwaarden mogelijk te maken zal er in de toekomst meer gekeken moeten worden naar de ontwikkeling van robots, in plaats van computers. Dit hoeven echter geen robots te zijn zoals we die nu kennen uit films, met een menselijke uitingsvorm. Het gaat hoofdzakelijk om verschillende sensorische ‘voelsprieten’, zoals bijvoorbeeld een camera. Veel AI-onderzoek richt zich daarnaast te veel op het ontwikkelen van algoritmen zonder de juiste context. Vaak worden algoritmen getraind in een gesimuleerde omgeving. Het ‘ontwikkelen in een vacuüm’ kan leiden tot algoritmen die weliswaar nauwkeurig zijn, maar niet de juiste problemen aanpakken ‘in de echte wereld’.

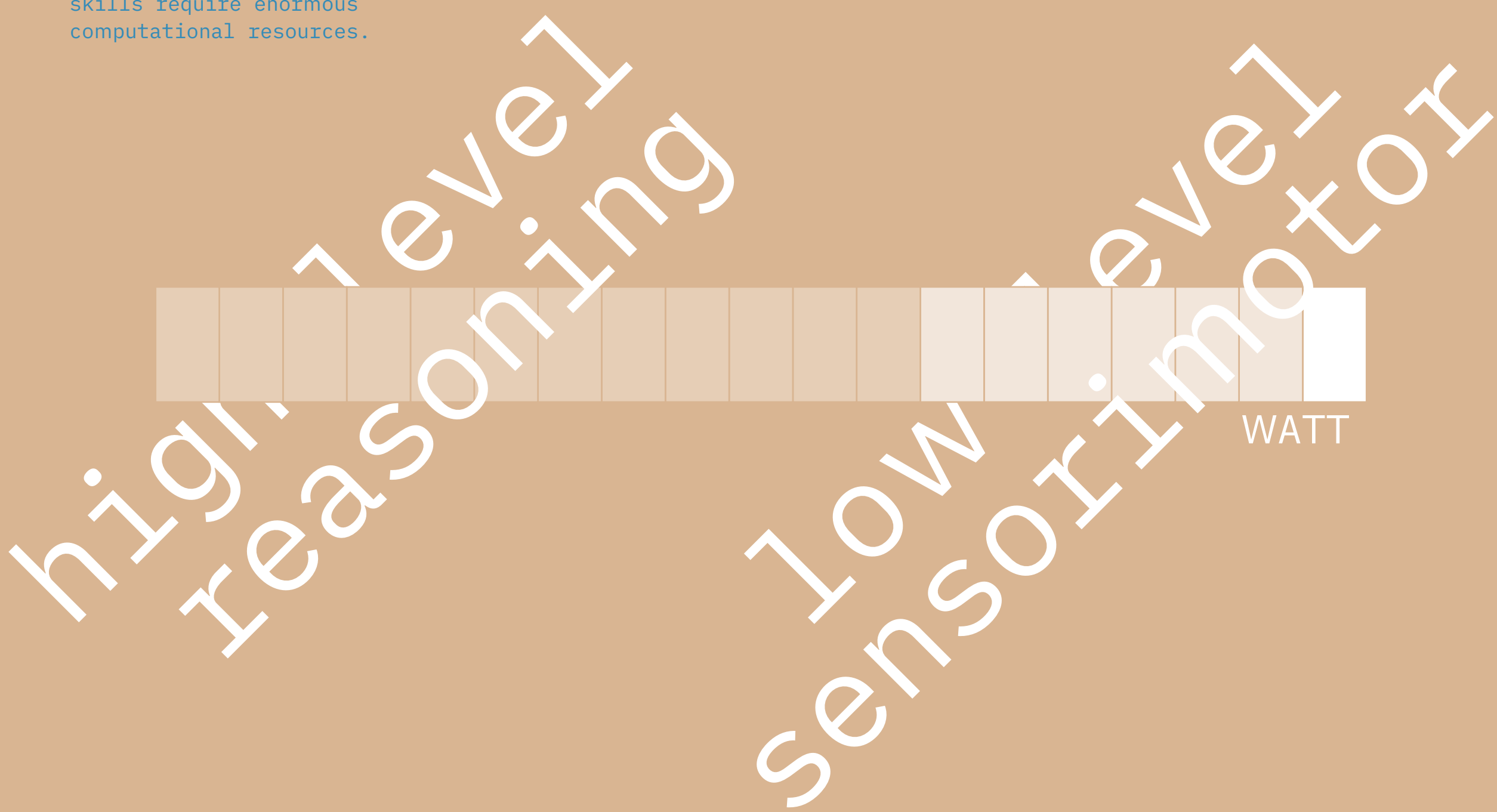
De wetmatigheden van intelligentie

Je kunt je afvragen waarom we überhaupt menselijke intelligentie nastreven. Het menselijk brein is qua structuur bijna het tegenovergestelde van een computer: complex en holistisch versus gestructureerd en binair. Een mens kan goed kijken, maar niet zo goed redeneren. Een computer kan goed redeneren, maar niet zo goed kijken. Juist wat wij intuïtief doen en ons weinig energie kost, is voor een computer complex en kost veel energie. En vice versa. Dit wordt ook wel *Moravec’s paradox* genoemd:

» *High-level reasoning requires very little computation, but low-level sensorimotor skills require enormous computational resources.* «
-- Hans Moravec, 1988

High-level reasoning requires very little computation, but low-level sensorimotor skills require enormous computational resources.

-- Hans Moravec, 1988



De huidige neurale netwerken ↴ zijn in verhouding tot het menselijk brein nog enorm klein. Mensen en computers maken daarbij hele andere fouten, bijvoorbeeld bij beeldherkenning. Zo kan een computer compleet van slag raken wanneer het aantal pixels wordt veranderd, terwijl dit voor mensen geen problemen zou opleveren. Tegelijkertijd trappen mensen in allerlei optische illusies, die een computer nooit van de wijs zouden brengen. We verleggen daarbij continu wat we intelligent vinden. Zodra iets werkt noemen we het geen AI meer. Het feit dat een navigatiesysteem ons door de stad kan leiden vinden we niet bijzonder meer. Als je daar een sticker 'nu met AI' op plakt, dan zouden mensen dat raar vinden (ook al is het feitelijk wel een vorm van AI).

We proberen al vanaf de jaren '50 menselijke intelligentie na te bootsen met computers en tegelijkertijd verleggen we steeds de horizon van wat we intelligent vinden. Een interessante paradox, maar wellicht ook een overbodige. De structuur en capaciteiten van computers zijn dermate anders dan die van ons brein, dat het soort intelligentie ook compleet anders is. De Turing test is vanuit dat perspectief de verkeerde meetlat, het is te antropomorfisch. Een zelfrijdende auto zou niet als de mens moeten rijden, de context moet juist meeveranderen. Net zoals de impact van de eerste auto's op de openbare weg niet te vergelijken is met een paard en wagen. Stel je voor dat de verkeersregels hier niet op aangepast waren. Grappig detail is dat de capaciteit van een motor wel nog steeds wordt uitgedrukt in paardenkracht (PK).

» The question of whether machines can think is about as relevant as the question of whether submarines can swim. «

-- Edsger W. Dijkstra, 1984

In plaats van het mechanisme van intelligentie in het brein te kopiëren (en de biologie die daarmee gepaard gaat) lijkt het logischer om de fysieke equivalenten

ervan zien te vinden, oftewel parameters van intelligentie, en deze proberen te optimaliseren. Net als bij een vliegtuig. We hebben niet letterlijk de vliegtechniek van een vogel gekopieerd, maar gekeken naar de krachten van vliegen, namelijk lift, gewicht, stuwkracht en slepende kracht. Pas toen mensen het idee loslieten dat we de vliegtechniek van een vogel moesten kopiëren (en niet meer met grote vleugels probeerden te fladderen) zijn er grote stappen gezet. De vraag is in hoeverre vergelijkbare krachten ook voor intelligentie te formuleren zijn.

Een formule voor intelligentie

Om de wetmatigheden in kaart te kunnen brengen is het belangrijk om naar de oorsprong van intelligentie te kijken. Biologische evolutie stelt een organisme in staat om te overleven door aangepast te blijven aan de veranderende omgeving. Het *timeframe* van deze verandering moet echter wel langer zijn dan de voortplantingstijd, anders kan het organisme zich niet aanpassen en dus niet overleven. Harvard professor Steven Pinker noemt het gat in deze aanpassingstijd in zijn onderzoek uit 2010 ook wel de *cognitive adaptive* niche. Om te kunnen anticiperen op snelle veranderingen zijn dus andere, meer cognitieve mechanismen nodig, zoals een voorstellingsvermogen. Mentale simulaties gaan namelijk sneller dan de realiteit, waardoor het mogelijk is om vooruit te kunnen denken. Deze cognitieve vermogens maken het mogelijk om toekomstige veranderingen in de omgeving voor te kunnen stellen en hierop te anticiperen. Door interne cognitieve simulaties kunnen keuzes gemaakt worden die de *future freedom of action*, oftewel de toekomstige vrijheid van handelen, maximaliseren. Intelligentie is in de basis dus een overlevingsmechanisme.

» Wie niet sterk is, moet slim zijn. «

Onderzoeker Alex Wissner-Gross stelt in zijn TED Talk uit 2013 de vraag of er een formule voor intelligentie

is. Hij gelooft van wel. Hij noemt het zelfs de meest vergelijkbare variant van E = mc² voor intelligentie.

$$F = T \nabla S\tau$$

Binnen deze vergelijking wordt intelligentie aangeduid als een *force* (F), die gedreven wordt om de toekomstige vrijheid van handelen te maximaliseren. Dit gebeurt met enige kracht (T) met de verscheidenheid van mogelijke toegankelijke toekomsten (S) in relatie tot de toekomstige tijdshorizon, tau (τ). Kort gezegd, intelligentie houdt er niet van om opgesloten te zitten. Het probeert de opties zoveel mogelijk open te houden. Intelligentie moet volgens Wissner-Gross gezien worden als een fysiek proces dat de toekomstige vrijheid van handelen probeert te maximaliseren en beperkingen in de eigen toekomst voorkomt.

Deze formule borduurt voort op het inzicht dat er een relatie bestaat tussen intelligentie en wanorde. Uit kosmologisch onderzoek blijkt dat een universum dat meer wanorde produceert betere condities creëert voor het bestaan van intelligente soorten. Om dit te onderzoeken ontwikkelde Wissner-Gross ‘Entropica’ (afgeleid van *entropy*, wat wanorde betekent). De software is zo ontworpen dat het de ontwikkeling van lange termijn wanorde stimuleert in elk systeem waarin het geplaatst wordt. Met deze insteek heeft de software geleerd om spellen te spelen, te handelen op de beurs en is het zelfs geslaagd voor verschillende intelligentietesten voor dieren, zónder hiervoor geprogrammeerd te zijn. In dat opzicht kun je stellen dat intelligentie voortkomt uit wanorde. Het impliceert echter ook dat onze angst dat machines ons gaan overheersen weliswaar niet klopt, maar niet volledig ongegrond is. Je kunt namelijk stellen dat machines niet eerst intelligent hoeven te zijn, voordat ze proberen controle over te nemen. De urgentie om controle te hebben over de mogelijke uitkomsten blijkt namelijk veel fundamenteler te zijn dan intelligentie.

» AI is not a fake intelligence, it is the next step in evolution. «

-- Justin Stanwix, Nanotronics

Geautomatiseerde besluitvorming

Besluitvorming is een vak apart. Verschillende vakgebieden houden zich bezig met besluitvorming en elk vakgebied heeft daarbij zijn eigen benadering. We maken voor deze toekomstverkenning onderscheid tussen drie verschillende benaderingen.

> *Cognitieve benadering:* Bij deze benadering ligt de focus voornamelijk op de mechanismen van besluitvorming. Een belangrijke vraag binnen deze benadering is op welke wijze oordeelsvorming werkt en op welke wijze mensen keuzes maken.

> *Bestuurskundige benadering:* Binnen deze benadering wordt besluitvorming voornamelijk gezien als onderdeel van beleidsvorming. Hierbij is er aandacht voor de gehele cyclus van totstandkoming tot implementatie.

> *Economische benadering:* Bij deze benadering ligt de focus voornamelijk op de strategische keuzes die in het proces gemaakt worden en de belangen die verschillende stakeholders hierin hebben. Binnen deze benadering wordt geprobeerd om besluitvorming te modeleren.

We willen graag geloven dat wij mensen rationele wezens zijn. Dat de keuzes die we maken en de besluiten die we nemen weloverwogen zijn, gebaseerd op kennis en feiten. Maar het besluitvormingsproces is grillig. Beeldvorming en ambities spelen naast feitelijke kennis een belangrijke rol. Deze componenten zijn niet alleen subjectief, maar ook erg veranderlijk. We laten ons continu beïnvloeden door irrationele componenten. Veel besluiten worden dan ook onbewust, bijna automatisch, gemaakt. Denk bijvoorbeeld aan sollicitatiegesprekken. Je weet intuïtief eigenlijk al heel snel of je iemand gaat aannemen of niet (Systeem 1 ↴). In de rest van het gesprek ben je voornamelijk bezig met de rechtvaardiging van het gevoel (Systeem 2 ↴). Daarbij komt dat de uitleg van onze keuzes achteraf helemaal niet zo goed te herleiden is als we onszelf doen geloven. AI ↴ daarentegen, zou keuzes veel meer objectief kunnen maken. Een machine handelt immers rationeel. Maar is dat wel zo? We kijken daarom eerst naar de cognitieve benadering en zetten deze per aspect af tegen de geautomatiseerde besluitvorming van AI.

Het irrationele brein

Een groot deel van ons gedrag en de keuzes die we daarbij maken worden gedreven door instinctieve en emotionele processen in ons brein. Aangezien we evolutionair zijn geprogrammeerd om zoveel mogelijk energie te besparen (en het dus ook logisch is dat we Systeem 1 het meeste werk laten doen) sluipen er wel wat ‘denkfouten’ in onze informatieverwerking. Deze denkfouten worden ook wel cognitive biases genoemd. Zo hebben we sterk de neiging om bewijs te zien dat onze eigen gedachten bevestigt *(confirmation bias)*, hebben we een voorkeur voor mensen die we tot dezelfde groep rekenen als wijzelf (ingroup-outgroupbias) en hebben we de neiging om álle karaktereigenschappen van een persoon of organisatie te waarderen *(halo-effect)*.

Een andere interessante bias is dat we een afkeer hebben van verlies; we reageren sterker op verliezen dan op winsten. Deze bias wordt ondersteund door de prospect theorie van Kahneman uit de jaren ’70. Deze theorie gaat over de wijze waarop mensen besluiten nemen onder risico en met hoge onzekerheid. Mensen blijken inderdaad niet alles keurig op een rijtje te zetten om te bekijken welke optie de hoogste waarde oplevert, maar mensen gaan uit van een referentiepunt; levert dit besluit winst dan wel verlies op. Dat bepaalt of ze risico’s willen nemen of niet. In het geval van winst nemen mensen voorzichtige, risicomijdende besluiten. Wanneer geconfronteerd met verliezen nemen mensen riskantere besluiten. Vaak gaat het om *split second decisions* en zijn dus meer instinctief dan rationeel.

AI gone wrong

Niet alleen menselijke intelligentie, maar ook AI heeft last van vooroordelen. De output van algoritmen ↴ kan namelijk *gender* en *race biases* bevatten. De verklaring ervan is simpel. Wanneer de input niet zuiver is, dan is de output dat ook niet.

» *Garbage In, Garbage Out.* «

Een bekend voorbeeld hiervan is de Twitter *chatbot* van Microsoft, genaamd Tay. Volgens Microsoft zou de chatbot steeds slimmer worden naarmate je er meer mee zou chatten. Binnen 24 uur na lancering werd Tay echter al offline gehaald, omdat het aan de lopende band racistische uitspraken deed. De chats die mensen met Tay voerden waren niet zo keurig als Microsoft gehoopt had. Nu is dit nog een redelijk naïef voorbeeld. Als je wel eens op social media platforms komt dan had je dit immers kunnen voorspellen. Het wordt echter schrijnender wanneer het aankomt op bijvoorbeeld krediet-beoordelingen, fraudedetectie of sollicitatieprocessen. Zo heeft Amazon in 2014 een AI-applicatie ontwikkeld om sollicitanten te evalueren om zo tot een selectie van de beste kandidaten te komen. Pas een jaar later kwamen ze erachter dat deze seksistisch was. Het probleem was dat het algoritme werd getraind met data van sollicitanten die de afgelopen 10 jaar bij Amazon gesolliciteerd hadden. Aangezien dit in de *techbranche* voornamelijk mannen zijn, kreeg het algoritme een voorkeur voor mannelijke kandidaten. In 2017 hebben ze de ontwikkeling van de applicatie stopgezet.

Bij dergelijke toepassingen worden vooroordelen – en daarmee digitale discriminatie – juist versterkt. Het probleem is dat data niet divers genoeg is. De vraag daarbij is of er voldoende data beschikbaar is van de meer kwetsbare groepen. Zo worden dialecten door spraak-herkenningssoftware nog onvoldoende herkend. Opvallend hierbij is dat ondanks het feit dat we al heel lang weten dat mensen bevooroordeeld zijn, we alsnog historische data gebruiken binnen gevoelige domeinen zoals de rechtspraak. In Amerika is veelvuldig gebruik gemaakt van software om te voorspellen hoe groot de kans is dat een veroordeelde opnieuw de fout in gaat. Uit onderzoek van ProPublica in 2016 blijkt dat deze software bevoor-oordeeld is in het nadeel van mensen met een donkere huidskleur. Biases van algoritmen worden dus veroorzaakt door biases van mensen.

Uitlegbaarheid

Mensen hebben behoefte aan een uitleg wanneer een bepaald besluit is genomen. Deze behoefte is goed zichtbaar in de zogenaamde ‘loterij paradox’.

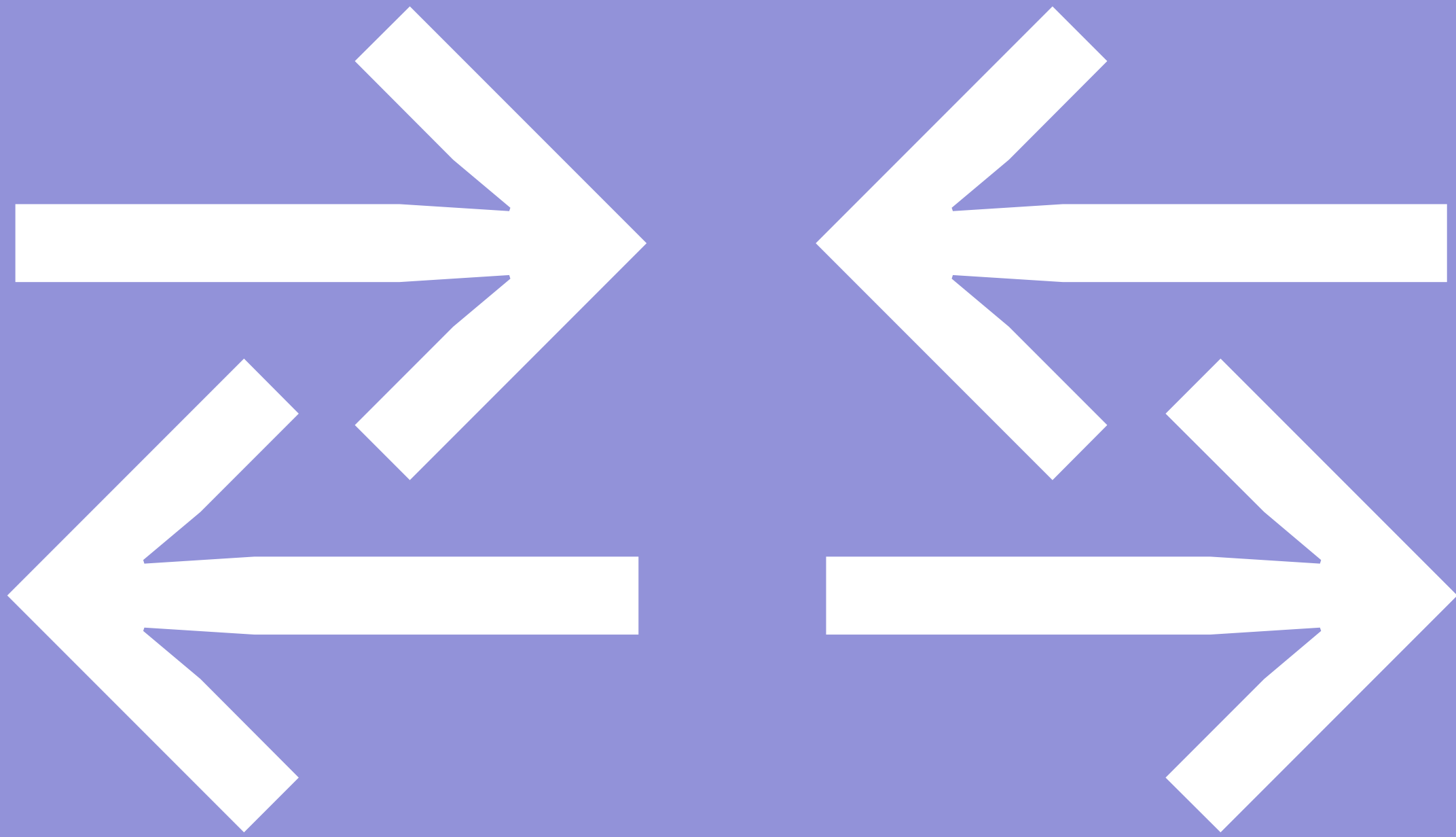
Wetenschappers kunnen meedingen naar beurzen door een onderzoeksvoorstel in te dienen. Wanneer je ze vraagt of ze liever willen dat een jury de prijzen uitdeelt, of dat het via loting verloopt, dan blijkt dat ze bijna unaniem aangeven dat ze liever een jury hebben. Via loting ontbreekt namelijk een uitleg. De paradox hierbij is dat ze de beoordeling van de jury ook als ‘loterij’ bestempelen.

Ondanks het feit dat de besluitvorming van mensen dus niet altijd volledig objectief is, wordt deze wel verkozen boven een random proces van loting. Er moet hierbij onderscheid gemaakt worden tussen transpa-rantie, de uitleg en uitlegbaarheid. Transparantie gaat voornamelijk over het proces en de opgestelde criteria vooraf, terwijl de uitleg en uitlegbaarheid over de toelichting op en herleidbaarheid van het besluit achteraf gaan. Bijvoorbeeld bij de jurybeoordeling.

> *Transparantie:* de beoordeling vindt plaats door middel van een zorgvuldig samengestelde expertjury. Zij scoren eerst individueel alle binnengekomen voorstellen. Op basis van deze scores wordt een top 5 samengesteld, waaruit op basis van een juryoverleg de winnaar wordt gekozen.

Garbage

Garbage



> *De uitleg:* de winnaar van deze aanvraagronde kwam als beste uit het juryoverleg naar voren en scoorde vooral hoog op creativiteit, mate van impact en mogelijkheden voor samenwerking. De jury was met name te spreken over de vernieuwende kijk op het onderzoeksgebied.

> *Uitlegbaarheid:* verschillende deelnemers blijken zich niet in de uitleg te kunnen vinden. Volgens hen is het winnende voorstel geen vernieuwende kijk op het onderzoeksgebied. In hun ogen is de keuze onvoldoende verdedigbaar en dus niet uitlegbaar.

Uitlegbaarheid is dus erg subjectief en contextafhankelijk, terwijl transparantie en de uitleg veel meer objectief zijn. Daarbij komt dat mensen het prettig vinden om de vinger te kunnen wijzen als ze het ergens niet mee eens zijn. *Accountability*, oftewel aansprakelijkheid, is dus een belangrijk onderdeel van het besluitvormingsproces. Dus als jij de beurs niet hebt gekregen, dan kun je de schuld altijd bij de jury neerleggen. Het voelt goed om iemand als de schuldige aan te kunnen wijzen. De schuld aan 'het mechanisme van loting' geven is wat lastig, aangezien deze geen gezicht heeft. Hetzelfde zie je in de politiek. Het is een soort afstrafmechanisme. Als een politieke partij slecht presteert kun je deze straffen door de volgende keer niet meer op ze te stemmen.

» Besluitvorming is rommelig. «
-- Barbara Vis, UU

Overigens zijn 'harde feiten' in dit proces niet altijd doorslaggevend. Uit een artikel uit 2017 van Barbara Vis, hoogleraar Politics & Governance, blijkt dat veel burgers niet alleen slecht, maar vaak ook verkeerd geïnformeerd zijn. Veel mensen schatten bijvoorbeeld het aantal terroristische aanslagen in Europa veel te hoog in. Het grote gevaar hierbij is dat deze mensen hun opvattingen doorgaans niet aanpassen wanneer ze geconfronteerd worden met de daadwerkelijke feiten. Mensen zijn geneigd om informatie te negeren wanneer deze niet overeenkomt met hun eigen opvattingen. Sterker nog, zelfs mensen die wéten dat hun opvattingen feitelijk onjuist zijn, zijn geneigd om bij hun standpunt te blijven wanneer ze vaak genoeg een leugen aangeboden krijgen.

Explainable AI

Besluitvormingsprocessen worden steeds vaker ondersteund door data. Algoritmen maken daarbij steeds vaker zélf de clusters ↴, zonder voorgeprogrammeerde labels. Het wordt hierdoor steeds lastiger voor mensen om te achterhalen op basis van welke data de uitkomsten gebaseerd zijn. AI wordt daarom nu nog vaak vergeleken met een black box. Uit onderzoek van de UvA naar geautomatiseerde besluitvorming door AI uit 2018 blijkt dat veel Nederlanders bezorgd zijn dat AI leidt tot manipulatie, risico's of onaanvaardbare resultaten. Alleen bij meer objectieve besluitvormingsprocessen, zoals de aanvraag van een hypotheek, werden er kansen gezien voor AI. Menselijke controle, menselijke waardigheid, eerlijkheid en nauwkeurigheid worden als belangrijke waarden genoemd bij het nadenken over besluitvorming door AI. Een belangrijke vraag is of het besluitvormingsproces in de toekomst voldoende transparant is en of de uitkomst voldoende uitlegbaar is. Er wordt daarom veel aandacht besteed aan Explainable AI. Cor Veenman, onderzoeker bij TNO, benoemt verschillende aspecten die van belang zijn voor transparantie in het geautomatiseerde besluitvormingsproces:

- Wat zijn de bedrijfsdoelstellingen en waarom?
- Van welke databases wordt gebruik gemaakt en waarom?
- Welke data is gebruikt en waarom?
- Welke modeleringstechnieken zijn gebruikt en waarom?
- Op welke wijze interpreteer je de uitkomsten en waarom?

Er lijkt echter sprake te zijn van een *trade off* tussen Explainable AI aan de ene kant, waarbij mensen vast

willen houden aan de mogelijke verklaringen die het model kan geven, en performance aan de andere kant, waarbij de nauwkeurigheid van nieuwe technieken zoals deep learning ↴ enorm hoog is, maar het vermogen van verklaarbaarheid deels moet worden opgegeven. Computers leren steeds meer intuïtief. Het is overigens niet zo dat een van de twee het alleenrecht heeft op de toekomst. Het is een 'knop' waar designers aan kunnen draaien en waar gebruikers hun wensen in aan kunnen geven.

» Er komt misschien wel een dag dat wij tegen AI zeggen 'verklaar jezelf', maar dat de AI zegt 'you wouldn't understand it anyway'. «

-- Maarten Stol, BrainCreators

Het is hierbij belangrijk om te bepalen wanneer je welke uitleg geeft en in welke context. *Explainability* is geen *silver bullet*, in sommige gevallen kan uitleg juist averechts werken. Een grote uitdaging daarbij is aansprakelijkheid. In theorie kunnen machines misschien wel betere besluiten maken dan mensen, maar het gevoel erbij klopt niet. Je kunt ze namelijk niet tot de verantwoording roepen.

Vertrouwen

Vertrouwen heeft te maken met onzekere situaties en is bij bijna alle interacties tussen twee partijen essentieel. In onzekere situaties blijkt vertrouwen zelfs een significante rol te spelen in het gedrag van mensen. In de basis gaat vertrouwen over het uit handen geven van kwetsbaarheden en het geloof dat de andere partij zich naar verwachting zal gedragen.

Vertrouwen hangt sterk samen met uitlegbaarheid. Je kunt stellen dat hoe meer vertrouwen je ergens in hebt, hoe minder uitleg je nodig hebt. Ik ken bijvoorbeeld niemand die de uitkomsten van zijn rekenmachine controleert. Het is in dat opzicht vergelijkbaar met de auto: ik begrijp niet hoe de motor werkt (en vind het ook niet nodig om te weten), maar ik heb wel een rijbewijs en kan er mee rijden. Ik vertrouw erop dat als een auto op de markt komt, dat deze aan voldoende kwaliteitsnormen en veiligheidstests onderworpen is. Ik heb overigens geen idee aan welke normen en tests, maar ik vertrouw erop dat het goed zit. Veel AI-technologieën staan nog in de kinderschoenen en staan aan het begin van het proces van vertrouwen. Wanneer er richtlijnen zijn opgesteld voor de beheersbaarheid van de technologie en er voldoende tests zijn uitgevoerd zal AI volgens deze zienswijze ook vertrouwd ↴ gaan worden. Net als alle andere technologieën die eerder gewantrouwd werden.

Algorithm aversion

Toch lijkt er met AI iets anders aan de hand te zijn. We accepteren al decennialang dat er door de introductie van de auto wereldwijd dagelijks meer dan 3000(!) mensen overlijden in het verkeer. Toch staat de wereld op zijn kop wanneer er als gevolg van een (semi)zelfrijdende auto één persoon overlijdt. In 2018 stopte Uber tests met de zelfrijdende auto's, nadat er een voetganger was aangereden en overleed in Arizona. Hieruit kun je concluderen dat we strenger zijn voor de technologie dan voor onszelf. Om vertrouwen te hebben in de technologie moeten we kwetsbaarheden uit handen geven aan een 'onzichtbare entiteit' die we niet tot de verantwoording kunnen roepen. Ook hier speelt het principe van aansprakelijkheid dus mee. Fouten maken is daarbij menselijk, vinden we zelf. Verschillende citaten en wijsheden ontlenen hieraan hun bestaansrecht.

» Wie niet in staat is een fout te maken, is tot niets in staat. «

-- Abraham Lincoln

Wanneer een algoritme een fout maakt is het echter een ander verhaal. Uit onderzoek van onder andere de University of Penssylvania uit 2014 blijkt dat wanneer mensen een algoritme een kleine en betekenisloze fout zien maken, dat de kans dan groot is dat het vertrouwen

volledig verloren gaat. Zelfs wanneer algoritmen aantoonbaar betere besluiten maken, dan nog geven mensen de voorkeur aan hun eigen onderbuikgevoel. Dit wordt door onderzoekers ook wel *algorithm aversion* genoemd. De onderbouwing van mensen is dat ze meer vertrouwen hebben in het lerend vermogen van mensen, dan in dat van machines. In een experiment werd gevraagd om op basis van toelatingsdata van een MBA-opleidingsprogramma te raden hoe goed studenten het gedaan hadden in het programma. De deelnemers aan het onderzoek werd verteld dat ze een klein geldbedrag konden winnen en dat ze ervoor konden kiezen om zelf een gok te wagen of het over te laten aan een algoritme. Zelfs de deelnemers die gezien hadden dat hun eigen voorspellingen bij eerdere tests slechter waren dan die van het algoritme, maakten alsnog de keuze om zelf de gok te wagen. Alleen wanneer deelnemers de uitkomsten van de voorspelling van het algoritme zelf naar boven of naar beneden konden bijstellen, bleek het vertrouwen omhoog te gaan en verkozen ze het algoritme boven hun eigen voorspelling.

Het lijkt erop dat mensen zich bedreigd voelen door technologie. Uit een onderzoek naar de competitie tussen mensen en robots uit 2019 blijkt dat mensen die verliezen van een algoritme sterk ontmoedigd worden om nog een keer te spelen. In het onderzoek werden deelnemers uitgedaagd om een wedstrijd aan te gaan tegen een algoritme in het tellen van de letter G en in een random reeks letters. Een goed antwoord leverde punten op, terwijl bij een foutief antwoord het scherm 10 seconden bevroren werd. Wanneer ze het algoritme zouden verslaan wonnen ze een geldbedrag. Dit bedrag werd af en toe aangepast om te kijken wat het effect was op de motivatie. Wat bleek, een hoger geldbedrag motiveerde deelnemers niet, het ontmoedigde ze juist. Deelnemers gaven aan zich gestrest te voelen wanneer ze het puntenaantal van het algoritme in hun ooghoek zagen stijgen.

Sympathie

Biases zorgen ervoor dat we eenvoudig beïnvloed worden in het besluitvormingsproces. Zo zijn we eerder bereid om toe te geven aan het verzoek van iemand die we sympathiek vinden. Sympathie, oftewel *liking*, is een van de zes beïnvloedingsmechanismen die Robert Cialdini beschrijft in zijn boek Influence: The Psychology of Persuasion uit 1984. Overeenkomstig met de ingroup-outgroupbias vinden we voornamelijk mensen sympathiek waar we veel overeenkomsten mee hebben. Denk hierbij aan mensen met een vergelijkbare achtergrond en interesse. Ook aantrekkelijke mensen vinden we sympathieker. We zijn geneigd om aantrekkelijke mensen automatisch allerlei positieve eigenschappen toe te schrijven die ze niet per se bezitten, zoals talent, vriendelijkheid, eerlijkheid en intelligentie. Een typisch geval van het halo-effect.

Vertrouwelijkheid wekt tevens een gevoel van sympathie op. Dit kan ontstaan door herhaaldelijk contact, maar ook door herkenning. We voelen ons bijvoorbeeld meer verbonden met iemand die dezelfde naam heeft als wijzelf. Ook mensen met een fysieke gelijkenis zijn we geneigd om sympathieker te vinden en vertrouwen we sneller.

How do you like me now?

De vraag is of we ons ooit daadwerkelijk kunnen herkennen in machines. De zelfrijdende auto heeft de potentie om het aantal ongelukken op de weg te verlagen. Maar om deze potentie te kunnen benutten moeten mensen wel in willen stappen en de controle uit handen durven geven. Het vertrouwen in de zelfrijdende auto blijft echter nog enorm laag. Uit jaarlijks onderzoek van AAA, een Amerikaanse organisatie die zich inzet voor automobilisten en veiligere wegen en voertuigen, blijkt dat 71% van de Amerikanen in 2019 geen vertrouwen heeft in de zelfrijdende auto. Slechts 19% van de ondervraagden zou hun familie of kinderen met een zelfrijdende auto de weg op sturen.

De vraag is onder welke voorwaarden mensen de technologie gaan omarmen. Frank Verberne, destijds

verbonden aan de TU Eindhoven, schreef er in 2015 een proefschrift over: Trusting a virtual driver; similarity as a trust cue. Zijn hypothese is dat het principe van gelijkenis (dat bij mensen sympathie en vertrouwen verhoogt) ook voor technologie zou moeten werken. We geven onze auto's immers namen en worden boos op een computer als deze niet wil opstarten. Verberne creëerde daarom een digitale avatar met overeenkomstige kenmerken van de proefpersoon en maakte deze op een scherm op het dashboard zichtbaar. Denk bijvoorbeeld aan een vergelijkbaar uiterlijk en een vergelijkbare manier van bewegen en communiceren. Een 'digitale chauffeur' blijkt het vertrouwen in de zelfrijdende auto inderdaad te verhogen. Mensen hebben het gevoel dat de auto bestuurd wordt door deze virtuele chauffeur, waarmee we ons vervolgens kunnen vereenzelvigen. Hierbij is de juiste balans echter wel cruciaal; wanneer de avatar teveel op de proefpersoon leek ging het effect verloren.

De vermenselijking van machines is op meerdere fronten zichtbaar. Zo bouwen we robots naar onze gelijkenis, geven we onze virtuele assistenten namen en een menselijke stem en is er zelfs een *lag* ingebouwd in chatbots. Deze vertraging in de beantwoording van de chatbots moet ervoor zorgen dat de gesprekken natuurlijker aanvoelen. We worden op deze manier wel een beetje voor de gek gehouden, maar waarschijnlijk accepteren we het met liefde. Onze intiemere relatie met technologie vormt een dankbare vorm van inspiratie voor filmmakers. Zo zien we in de film Her uit 2013 hoe de hoofdpersoon verliefd wordt op het besturingssysteem van zijn computer.

De toepasbaarheid van AI in besluitvorming

Besluitvorming is een grillig proces. Zo zijn er verschillende irrationele componenten van invloed op het menselijke besluitvormingsproces. Maar ook algoritmen zijn niet vrij van biases en hebben moeite met uitlegbaarheid. Voordat AI kan worden ingezet voor besluitvorming is het daarom van belang om maatregelen te

nemen die de transparantie van het proces en de uitlegbaarheid van de uitkomst waarborgen. Hierbij dient goed nagedacht te worden over de tradeoff tussen performance en explainability.

In de huidige discussies omtrent de toepasbaarheid van AI bij besluitvormingsprocessen wordt er echter onvoldoende onderscheid gemaakt tussen de verschillende typen besluiten en de impact die de besluiten hebben op de betrokkenen. Het maakt nogal een verschil of het gaat om een aanbeveling voor een film, een diagnose op basis van longfoto's in het ziekenhuis of een advies ten aanzien van een bedrijfsovername. De mate van impact is dus onder andere afhankelijk van de mogelijke risico's. Naast de cognitieve benadering, moeten we besluitvorming dus ook vanuit de bestuurlijke en economische benadering bekijken.

De impact van besluitvorming

Om te kunnen bepalen wat de toepasbaarheid is van geautomatiseerde besluitvorming, zijn er verschillende factoren die in overweging genomen moeten worden. Deze factoren bepalen de impact van de besluitvorming.

1: Mate van subjectiviteit (hoog/laag):

Het is belangrijk om te bepalen in welke mate er overeenstemming is, oftewel consensus, over de juistheid van de uitkomst van het proces. In sommige gevallen is er sprake van een algemeen geldende opvatting en delen de meeste mensen dezelfde mening. In andere gevallen bestaat er geen algemeen geldende opvatting en zijn de opvattingen veel meer subjectief.

2: Mate van onzekerheid (hoog/laag):

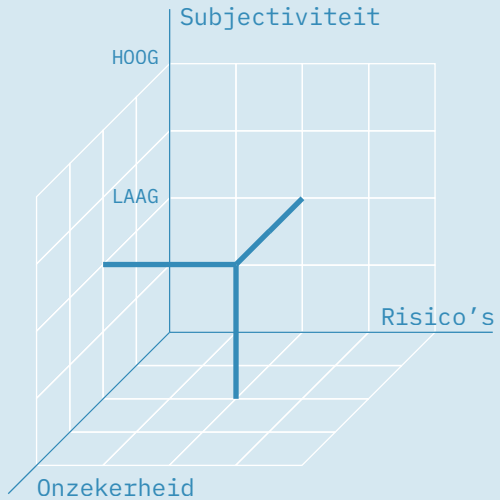
In het geval van geautomatiseerde besluitvorming is de mate van onzekerheid sterk afhankelijk van de beschikbare data. In sommige gevallen is de data onvolledig en kunnen niet alle aspecten direct afgeleid worden uit de voorhanden zijnde data. De frequentie waarop het besluitvormingsproces zich voordoet heeft hier tevens invloed op. Wanneer de frequentie laag is, is er vaak te weinig

data beschikbaar om gegronde uitspraken te kunnen doen. Dit verhoogt de onzekerheid van de output.

3: Mate van risico's (hoog/laag)
Het besluitvormingsproces kan verschillende risico's met zich meebrengen. Vaak worden risico's door organisaties uitgedrukt in financiële kosten. Het gaat dan niet zozeer over de kosten van het besluitvormingsproces zelf, maar om de financiële gevolgen van een foutieve beslissing. Daarnaast zijn er ook andere risico's mogelijk. Denk bijvoorbeeld aan reputatieschade van de organisatie of fysiek letsel van betrokkenen.

To automate or not to automate? That's the question

De scores op de verschillende factoren bepalen de mate van toepasbaarheid van AI bij besluitvormingsprocessen. Hierdoor ontstaat het volgende assenstelsel, dat we de *Automated Decision Making Index* (ADMI) genoemd hebben.



De Automated Decision Making Index (ADMI):
voorbeeld van Subjectiviteit laag/Onzekerheid laag/Risico's laag

Er zijn op basis van de verschillende factoren acht mogelijke combinaties:

- Subjectiviteit laag / Onzekerheid laag / Risico's laag
- Subjectiviteit hoog / Onzekerheid laag / Risico's laag
- Subjectiviteit laag / Onzekerheid hoog / Risico's laag
- Subjectiviteit laag / Onzekerheid laag / Risico's hoog
- Subjectiviteit hoog / Onzekerheid hoog / Risico's laag
- Subjectiviteit hoog / Onzekerheid laag / Risico's hoog
- Subjectiviteit laag / Onzekerheid hoog / Risico's hoog
- Subjectiviteit hoog / Onzekerheid hoog / Risico's hoog

Deze combinaties beschrijven de verschillende typen besluitvorming en bepalen in welke mate AI toepasbaar is bij besluitvormingsprocessen. Er zijn in de basis drie mogelijke opties:

> *Automatiseren:* AI kan het besluitvormingsproces volledig autonoom doorlopen.
> *Ondersteunen:* AI kan mensen in het besluitvormingsproces ondersteunen in de vorm van advies.
> *Niet automatiseren:* AI kan geen rol spelen in het besluitvormingsproces.

Momenteel wordt AI voornamelijk operationeel ingezet. Denk aan slimme AI-toepassingen voor onder andere spoordetectiesystemen en medische diagnoses. Maar ook de aanbevelingen van je streamingsplatform en het automatisch inhalen van je zelfrijdende auto zijn allemaal voorbeelden van operationele taken. Dingen die zich goed lenen voor *trial and error*, waar veel data is en je gemakkelijk kunt experimenteren. Een voorbeeld is Netflix, waarbij de subjectiviteit van de aanbevelingen wellicht hoog is, maar welke door de lage onzekerheid en risico's heel goed te automatiseren is. Bij medische diagnoses daarentegen is de subjectiviteit en onzekerheid laag, maar de risico's dermate hoog dat het waarschijnlijk verstandig is om deze niet volledig te automatiseren, maar door AI te laten ondersteunen.

Bij de zelfrijdende auto wordt vaak het *trolley problem* aangehaald om te laten zien hoe subjectief de keuzes van de zelfrijdende auto zijn. De vraag is dan wat een auto moet doen wanneer er bijvoorbeeld een groep mensen bij een zebrapad oversteekt en de auto niet meer op tijd kan remmen. Er zijn dan theoretisch gezien twee opties: doorrijden om de inzettende te sparen of uitwijken om de overstekende mensen te sparen. En maakt het dan nog uit wie er oversteken en wie er in de auto zitten? MIT ontwikkelde hiervoor de Moral Machine om te kijken wat je zelf in zo'n situatie zou doen. Toch kun je je afvragen of dit de ontwikkeling van de zelfrijdende auto moet tegenhouden. Hoe vaak komt dit namelijk in de praktijk daadwerkelijk voor? En verhoudt zich dit tot het aantal ongelukken dat er voorkomen kan worden? De computer zal in de basis de juiste keuzes maken, omdat we het er over het algemeen wel over eens zijn wat er goed en fout is in het verkeer. Het is daarom van belang om de verschillende factoren goed te overwegen en tegen elkaar af te zetten. Hierbij dient tevens rekening gehouden te worden met de voordelen die de inzet van AI bij besluitvormingsprocessen kan hebben, zoals het tegengaan van files en het verminderen van het aantal verkeersdoden.

De vraagt blijft momenteel nog of je algoritmen in de toekomst kunt inzetten bij strategische vraagstukken. Denk bijvoorbeeld aan een bedrijfsovername. De uitdaging zit in het feit dat er relatief gezien weinig data beschikbaar is. Er zijn dus veel onzekerheden en de risico's zijn hoog. Voorlopig zou je dus kunnen concluderen dat het bij strategische vraagstukken waarschijnlijk bij aanbevelingen door AI blijft.

Voordelen en beperkingen

In de zoektocht naar het benaderen van human level AI ↴ komen we steeds meer te weten over onszelf. De technologie vormt als het ware een spiegel. AI ↴ maakt het daarbij mogelijk dat we experimenten kunnen uitvoeren omtrent intelligentie ↴ zonder dat we daarvoor mensen hoeven in te zetten. We kunnen dingen veel makkelijker uitproberen. Zo kunnen we onderdelen in een systeem plaatsen én ze er weer uithalen om te bekijken wat de effecten zijn. Bij mensen gaat dat niet zo gemakkelijk. Tegelijkertijd doet het ons beseffen dat we nog weinig weten over hoe intelligentie precies in ons brein werkt.

Hetzelfde geldt voor de manier waarop wij besluiten nemen. Het is geen eenduidig rationeel proces, het is irrationeel en veranderlijk. Wanneer we het afzetten tegen geautomatiseerde besluitvorming hebben we echter de neiging om strenger ↴ te zijn voor de technologie, dan voor onszelf. Immers, fouten maken is menselijk. De biases ↴ in machines daarentegen vinden we onacceptabel. Toch stoppen we de vooroordelen er in veel gevallen zelf in.

» AI is een representatie van de werkelijkheid, niet van de wenselijkheid. «

Het gaat hoe dan ook om een kunstmatige vorm van intelligentie. Dus ook zonder de directe vergelijking met de werking van het menselijk brein te maken, is het interessant om te kijken naar de voordelen en beperkingen van de technologie.

Voordelen

AI wordt nu al veelvuldig ingezet voor het verhogen van de efficiëntie en het verbeteren van voorspellingen en klantrelaties. Het helpt ons om betere besluiten te nemen door de verwerking van grote hoeveelheden informatie in een korte tijd. Ook wanneer we general AI ↴

nooit zullen realiseren, kunnen we alsnog enorm veel bereiken met narrow AI ↴. De industrie loopt nu nog 10 jaar achter op wat er wetenschappelijk mogelijk is. We kunnen AI in dat opzicht vergelijken met de ruimtevaart.

» De vraag of we ooit daadwerkelijk op Mars gaan wonen is niet zo relevant. Maar de verkenning heeft ons wel enorm veel opgeleverd. «

Verschillende innovaties die we in het aardse leven gebruiken komen voort uit ruimtevaartprogramma's. Van het schokabsorberende effect van je Nike Air-schoenen en de antiaanbaklaag van je Tefal-pan, tot satelliet-communicatie en GPS-systemen van je mobiele telefoon. Het zijn allemaal spin-offs van technologieën die voor de ruimtevaart bedacht zijn. Zo zou de zoektocht naar human-level AI verschillende bruikbare innovaties kunnen voortbrengen, zonder dat human-level AI ooit bereikt wordt.

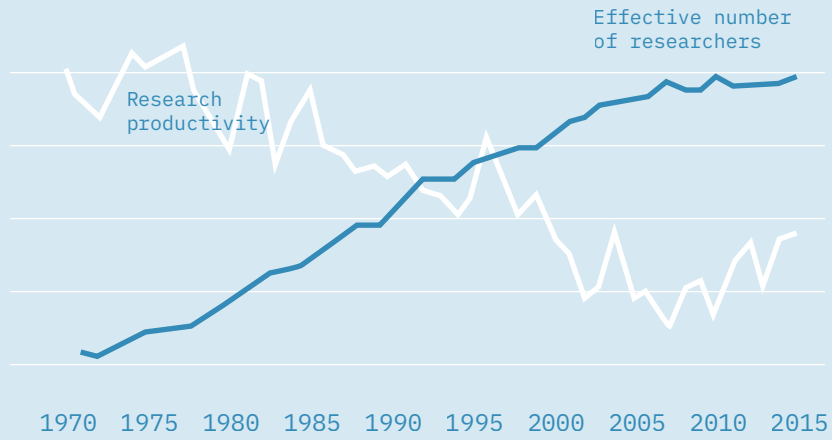
1 – 0 voor de technologie

Een groot voordeel van machine learning ↴-technieken is dat één getraind algoritme in meerdere systemen geïntegreerd kan worden. Bijvoorbeeld bij de zelfrijdende auto. Je hoeft niet elke auto afzonderlijk te trainen ↴, ze kunnen slim opgeleverd worden. Ze leren daarbij van elkaar, waardoor het een cumulatief leereffect heeft. Bij menselijke chauffeurs moet ieder individu afzonderlijk getraind worden en rijlessen nemen. Een rijinstructeur kan daarbij maar één persoon tegelijkertijd in de praktijk trainen en de onderlinge ervaringen worden niet direct uitgewisseld.

AI wordt tijdens het leerproces daarbij niet moe en klaagt nooit. Zelfs niet wanneer de stof enorm eentonig is of oneindig veel lijkt te zijn. Integendeel, zo heeft een algoritme het juist graag. Zo kan machine learning ons in potentie enorm helpen bij het ontdekken van bijvoorbeeld nieuwe medicijnen. Het ontwikkelen van

medicijnen is nu nog een enorm tijdrovende en kostbare klus. Wetenschappers moeten oneindig veel combinaties van moleculen proberen om te achterhalen welke verbindingen zorgen voor stabiele moleculen die bruikbaar zijn voor medicijnen. De meeste pogingen falen. Er zijn wel 10^{60} mogelijkheden denkbaar. Een algoritme kan met een grote rekenkracht, vierentwintig uur per dag en zonder tegenzin combinaties proberen. Dat kan enorm veel tijd en geld besparen. Én innovatie een boost geven.

Algoritmen kunnen patronen herkennen die wetenschappers niet zouden zien. Enerzijds door de grote hoeveelheden data, maar ook omdat machines 'Anders denken ↴'. Algoritmen zouden bepaalde opties kunnen onderzoeken die wij als naïef zouden bestempelen, maar welke juist voor een doorbraak kunnen zorgen. Iets wat ook wel nodig blijkt te zijn. Uit onderzoek van Stanford University en MIT uit 2019 blijkt dat de productiviteit van onderzoek daalt. Met andere woorden, er zijn steeds meer onderzoekers nodig om tot vergelijkbare resultaten te komen. AI kan daarbij onderzoek opschalen en daarmee de productiviteit verhogen.



'New molecular drug entities'– Bron: Stanford University en MIT (2019)

Algoritmen worden daarnaast steeds beter in het herkennen van uitzonderingen. Wanneer een algoritme weet wat men verwacht te gaan zien, kan het dingen herkennen die daarvan afwijken. Zoals bijvoorbeeld bij ruimteverkenningen. Zo hebben onderzoekers van de University of Texas in 2019 twee nieuwe planeten ontdekt met behulp van deep learning-algoritmen. Hierbij werd gebruik gemaakt van 'oude data' uit archieven van NASA. Ook op het gebied van onder andere veiligheid en inspectie kan dit veel voordeel opleveren.

» We don't want to miss something just because we didn't know to look for it . «

-- Kiri L. Wagstaff, NASA

De technologie kan daarnaast op plaatsen komen waar het voor mensen gevaarlijk is, zoals in oorlogsgebieden, bij een brand of in een kerncentrale. Voor die laatste toepassing heeft Guszti Eiben, hoogleraar in AI, een technologische revolutie ontwikkeld. Of beter gezegd, een technologische evolutie. Hij is in staat om robots als het ware te laten 'voortplanten', zodat ze kunnen evolueren tot de meest ideale vorm. Twee afzonderlijke robots zoeken elkaar op en wisselen informatie uit. Deze combinatie vormt de blauwdruk voor een nieuwe robot, die met een 3D-printer geprint kan worden. Een robot die in staat moet zijn om verouderde kerncentrales te verkennen moet daarvoor door de meest onmogelijke ruimtes kunnen kruipen. Door artificiële evolutie ontstaan er vanzelf robots die daar het beste toe in staat zijn.

» We openen hiermee de weg naar een geheel nieuw type evolutie: die van zelfstandig opererende machines. «

-- Guszti Eiben

AI for the good

AI heeft de potentie om ingezet te worden om het leven van mensen niet alleen gemakkelijker, maar ook beter te maken. Zo kan AI ook voordat volledig autonome voertuigen hun weg vinden al bijdragen aan de verkeersveiligheid. Door middel van affective computing ↴ kan AI emoties van autobestuurders lezen en interpreteren. Wanneer een bestuurder bijvoorbeeld tekenen van vermoeidheid toont, dan kan het systeem de bestuurder hierop wijzen en adviseren om bijvoorbeeld een kop koffie te gaan drinken bij het eerstvolgende tankstation. Zo kan AI dus ook een bijdrage leveren door in de auto te kijken, in plaats van alleen erbuiten.

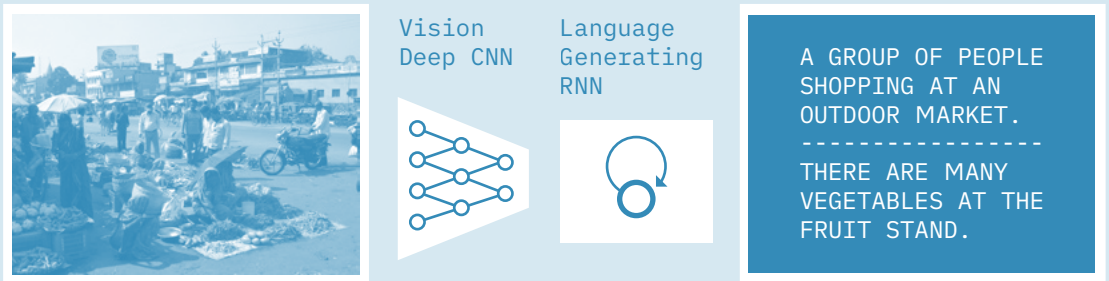
De inzet van *autonomous delivery robots* heeft op een campus van de George Mason University in de VS ervoor gezorgd dat studenten niet langer hun ontbijt overslaan. 88% van de studenten eet geen ontbijt, vaak als gevolg van een drukke ochtendplanning. De autonome bezorging draagt hierdoor bij aan een meer gebalanceerd voedingspatroon van de studenten. Dichterbij huis helpt het algoritme van de 'Slim Stampen' app scholieren bij het leren voor hun toetsen. Slim stampen is een adaptief overhoorprogramma dat zich aanpast aan de kennis en vaardigheden van de leerling. De applicatie zorgt ervoor dat een bepaald onderdeel precies op het juiste moment wordt aangeboden; net voordat een leerling het dreigt te vergeten. Wanneer een bepaald onderdeel te snel wordt aangeboden wordt het saai voor een leerling, maar wanneer het te laat wordt aangeboden vergeet de leerling het. Het systeem past de interne modellen na iedere test aan op basis van de correctheid van het gegeven antwoord en de snelheid waarop de vraag wordt beantwoord. Op deze wijze wordt het leerrendement sterk verhoogd.

AI kan ook bestuurders helpen om beter geïnformeerde besluiten te nemen. Zo kan het publieke opinies nauwkeurig in kaart brengen en een betere afspiegeling geven van wat mensen daadwerkelijk bezighoudt. De *wisdom of the crowd* kan tevens worden ingezet om

zwakke plekken in softwarebeveiliging te detecteren. Er is namelijk een plek waar bugs in systemen het allereerst gerapporteerd worden en 24 uur per dag geüpdatet wordt; Twitter. Onderzoek van Ohio State University uit 2019 toont aan dat het algoritme niet alleen in staat is om beveiligingsfouten te voorspellen, het blijkt tevens te kunnen voorspellen hoe hoog het risico is met een nauwkeurigheid van meer dan 80%.

Beperkingen

AI heeft de afgelopen decennia grote sprongen gemaakt en is tot dingen in staat die we tien jaar geleden niet voor mogelijk hielden. Zo kunnen computers met behulp van deep learning ↴ naast beelden herkennen, deze ook voorzien van een beschrijving.



Vereenvoudigde weergave van het proces waarbij een afbeelding door neurale netwerken wordt herkend en wordt beschreven

Het feit dat een AI-systeem een nooit eerder gezien beeld kan herkennen (op basis van gelabelde trainingsdata) en een beschrijving kan geven van de foto is weliswaar indrukwekkend, maar het betekent niet dat het systeem begrijpt ↴ wat het beeld representeert. Het systeem weet niet wat het betekent.

> *» The system only knows they are labeled as such. We have a tendency to interpret this beyond what is given. «*
>
> -- Eric Postma, Tilburg University

Afhankelijk van data

Een van de grootste beperkingen van AI is dat het afhankelijk is van grote hoeveelheden data. Bij deep learning wordt daarom ook wel gesproken over *data-hungry neural networks* ↴. Dit maakt dat de technologie niet goed is in randgevallen, waarbij weinig data beschikbaar is. Zo kan een zelfrijdende auto tijdens Halloween compleet van slag raken wanneer er een kind verkleed als spook de straat oversteekt. Vaak zijn datasets ook niet volledig. Uit onderzoek van het Georgia Institute of Technology uit 2019 blijkt dat zelfrijdende auto's minder nauwkeurig zijn bij het detecteren van voetgangers met een donkere huidskleur. Beeldherkenningssystemen zijn daarnaast heel gemakkelijk voor de gek te houden. Zo schijnt het *facial recognition system* van de iPhone X de gezichten van gebruikers niet te herkennen wanneer ze net wakker zijn geworden. Zorgelijker wordt het wanneer het om systemen gaat die worden ingezet bij het diagnosticeren van bijvoorbeeld huidkanker. Door een paar pixels te veranderen wordt de diagnose compleet anders. AI-systemen zijn daarbij vrij inflexibel. Ze zijn weliswaar in staat om zichzelf te leren schaken en grootmeesters te verslaan, maar wanneer je één spelregel verandert moet het algoritme helemaal opnieuw beginnen met leren.

Data heeft daarbij geen eeuwig leven (ook al denken we vaak van wel) Het is vergankelijk. De levensduur van een hard drive is beperkt en ook een datacenter is kwetsbaar: denk aan een aardbeving of vloedgolf. Daarnaast wordt er gewaarschuwd voor zonnestormen, die op aarde geomagnetische stralingen veroorzaken en de elektriciteits- en communicatienetwerken langdurig kunnen platleggen. De *cloud* lijkt soms ontastbaar, maar bedrijven als Google bouwen gigantische datacentra met zo'n 900.000 servers.

Verder zijn de formaten van de opslagmedia en de afspeelapparatuur vergankelijk. Evenals de software waarmee we bestanden maken en openen. Er zijn daarnaast ook gevaren vanuit cybercrime. Systemen kunnen gehackt worden en op deze wijze platgelegd worden. Doordat technologieën onderling steeds meer met elkaar interacteren wordt het risico dat de continuïteit van systemen in gevaar komt alsmaar groter. En als data de nieuwe olie is, dan zijn satellieten de nieuwe pijpleidingen. Naast satellieten zwerft er ook een hoop puin rondom de aarde, zoals raketonderdelen en afgedankte satellieten. In totaal meer dan 29 duizend stuks groter dan tien centimeter, die samen ruim acht miljoen kilo wegen. Deze toenemende *space waste* vormt een steeds realistischer gevaar voor ons dataverkeer. Wanneer deze in botsing raken met onze satellieten dan kan het zomaar zijn dat Facebook, WhatsApp en Instagram in een keer niet meer werken.

Computers zijn daarnaast enorme energieverbruikers. Zo kosten dertig Google *searches* net zoveel energie als het opwarmen van een liter water. Het menselijk brein verbruikt slechts twintig watt aan energie, terwijl supercomputers als IBM's Watson megawatts aan energie verbruiken (1 megawatt = 1.000.000 watt). De verwachting is dat de zelfrijdende auto meer energie gaat verbruiken voor het rekenwerk, dan voor het daadwerkelijk voortbewegen.

» Intelligentie per kilojoule zou een goede uitdrukking voor intelligentie kunnen zijn. «

-- Max Welling, UvA

Be careful what you wish for

In 540 voor Christus wenste Koning Midas dat alles wat hij aanraakte in goud zou veranderen. Hij stierf eenzaam en uitgehongerd, omdat ook zijn eten en geliefden in goud veranderden. Bij het formuleren van zijn doel dacht hij niet goed na over de gevolgen. Dit wordt ook wel het *Value Alignment Problem* (VAP) genoemd. Zo zou een zelfrijdende

auto bij de opdracht 'breng me zo snel mogelijk naar het vliegveld' met 300 kilometer per uur dwars door het weiland kunnen scheuren. Zo snel mogelijk betekent objectief gezien namelijk letterlijk zo snel mogelijk. Een machine kan zo doelgedreven zijn, dat de resultaten niet overeenkomen met wat we willen. Er zijn al ontwikkelingen gaande waarbij machines worden geprogrammeerd om *all means necessary* in te zetten om een doel te bereiken, dus bijvoorbeeld ook liegen. Zo is Facebook erin geslaagd een chatbot te ontwikkelen die geleerd heeft om te onderhandelen bij de aankoop van onder andere boeken. Op basis van menselijke conversaties heeft de chatbot het principe van misleiding weten te meesteren. Zo gaf de chatbot aan dat het interesse had in een bepaald object dat voor hem geen waarde had, om deze vervolgens uit te ruilen voor een betere deal op hetgeen het echt wilde hebben. Theoretisch gezien zou een intelligente machine die geprogrammeerd is om zoveel mogelijk paperclips te produceren álles in werking kunnen stellen om dat doel te bereiken. Nick Bostrom filosofeert in het boek Superintelligence ↴ dat de machine dan alles wat de productie belemmert, uit de weg zal ruimen. Zelfs de mens, want die draagt immers niet bij aan de productie van paperclips.

Je zou kunnen filosoferen dat een dermate intelligente machine dan zelf ook wel kan bedenken dat dit niet in lijn is met het doel, maar voor een algoritme is alle data in de basis gelijkwaardig. Dat levert bijvoorbeeld bij YouTube problemen op met ongewenste aanbevelingen. Wanneer mensen video's met racistische content bekijken en liken, dan zullen zij steeds meer vergelijkbare content getoond krijgen. De aanbevelingssystemen van YouTube zijn verantwoordelijk voor 70% van de totale tijd die gebruikers op de site doorbrengen. Dit leidt tot de zogenaamde *filter bubble*. Mensen krijgen dan geen informatie meer te zien die hun eigen standpunt tegenspreekt, waardoor mensen geïsoleerd raken in hun eigen 'luchtbel'. Dit leidt tot een eenzijdig wereldbeeld en kan polarisatie vergroten en een sterke invloed

hebben op bijvoorbeeld verkiezingen. Dit maakt dergelijke mediaplatforms enorm machtig.

> *» We embed our values in algorithms. «*
>
> -- Cathy O'Neil

Om dergelijke algoritmen te kunnen programmeren moet je bepalen wat een succesvolle uitkomst is. Maar wie bepaalt wat succesvol is? Bijvoorbeeld een algoritme die op basis van ingrediënten bepaalt wat je kunt koken: dan is de uitkomst afhankelijk van degene die bepaalt wat succes is. De moeder die wil dat haar kinderen groenten eten? De kinderen zouden ongetwijfeld iets anders als succes definiëren. Nu ligt de vraag wat een succesvolle uitkomst is nog in de handen van een zeer kleine groep mensen.

AI for the bad

De geschiedenis leert ons dat technologische innovaties niet alleen voor het goede ingezet worden. Denk bijvoorbeeld aan kernenergie. De potentie van AI kan dus ook misbruikt worden. Niet de machine zelf, maar de programmeur kan slechte intenties hebben. Uit onderzoek van Stanford University uit 2018 blijkt dat neurale netwerken in staat zijn om met een nauwkeurigheid van 91% de seksuele geaardheid van mensen af te lezen aan de hand van afbeeldingen van gezichten. Wat als dit wordt toegepast in een regime waar homoseksualiteit wordt afgekeurd? Met behulp van de technologie kan een totalitaire staat gecreëerd worden; *Big Brother is watching you*. Zo rolt China stap voor stap een 'sociaal kredietsysteem' uit. Chinese burgers krijgen op basis van hun gedrag een bepaalde score. Op basis van deze score kunnen mensen op een zwarte lijst worden geplaatst en allerlei (voor)rechten verliezen, zoals de mogelijkheid geld te lenen of naar het buitenland te reizen. In 2018 werden al 23 miljoen Chinezen verboden een trein- of vliegticket te kopen.

Wanneer je deze gedachte verder doorvoert kan het zogenaamde *singleton* principe ontstaan. Filosoof Nick Bostrom beschrijft een singleton als een hypothetische wereldorde waarin slechts één besluitvormingsorgaan bestaat. Mogelijke bevoegdheden zijn het vermogen om elke bedreiging (intern of extern) voor het eigen bestaan en suprematie te voorkomen, plus het vermogen om effectieve controle uit te oefenen over belangrijke kenmerken binnen het domein (zoals belastingheffing). De vraag blijft of een machine naast intelligente, ook wijze besluiten kan maken. Technologie heeft in principe geen moraal; ze geeft enkel de beste reactie gebaseerd op het verleden. En in hoeverre kan de technologie inschatten of een bedreiging daadwerkelijk reëel is? Het wordt steeds lastiger om echt en nep van elkaar te onderscheiden. Met behulp van AI kun je iemand dingen laten zeggen of doen die hij of zij in werkelijkheid nooit gezegd of gedaan heeft. Deze techniek wordt ook wel *deepfake* genoemd, een combinatie van de woorden deep learning en fake. Een bekend voorbeeld is de gemanipuleerde video van Barack Obama.

> *» We're entering an era in which our enemies can make anyone say anything at any point in time. «*
>
> -- Barack Obama's deepfake

Met behulp van de deepfake software kunnen video's beeld voor beeld geanalyseerd worden. Op deze wijze leert de software wat de omvang en vorm van een gezicht is, wat de verhouding is tussen de ogen, mond en neus, en hoe het gezicht beweegt wanneer er gepraat wordt. Ook stemmen kunnen worden nagemaakt. Op basis van dit model ontstaat een digitale versie van een persoon die je vervolgens alles kunt laten zeggen wat je wilt. Dergelijke software wordt steeds beter én steeds meer publiek toegankelijk.

Gebrek aan diversiteit

Niet alleen op het gebied van data, maar ook binnen de teams die aan AI-ontwikkelingen werken is er nog

“WE’RE ENTERING AN ERA IN WHICH OUR ENEMIES CAN MAKE ANYONE SAY ANYTHING AT ANY POINT IN TIME.”

-- Barack Obama’s deepfake

weinig sprake van diversiteit. In 2019 lanceerde Stanford University een instituut met als doel om onder andere het gebrek aan diversiteit in de industrie aan te pakken; The Institute for Human-Centered Artificial Intelligence (HAI). Van de 121 faculteitsleden die aanvankelijk werden aangekondigd, waren er meer dan 100 wit en het merendeel man. Dit wordt ook wel AI's *sea of dudes* of *white guy problem* genoemd.

In 2019 richtte Google een ethische raad op om feedback te krijgen op de AI-projecten van het bedrijf; *The Advanced Technology External Advisory Council* (ATEAC). Binnen vier dagen werd de raad opgeheven. Bijna duizend Google-medewerkers, academische onderzoekers en andere sleutelfiguren uit de tech-industrie hebben een brief ondertekend, waarin ze protesteerden tegen de samenstelling van de raad. Met name de aanstelling van Kay Coles James, president van de Heritage Foundation, werd met veel ongenoegen ontvangen. De Heritage Foundation is een denktank die onder andere heeft geprotesteerd tegen de bescherming van LGBTQ-rechten.

Volgens Rediet Abebe, computer science researcher bij Cornwell University en oprichtster van Black in AI dat zich inzet voor de inclusie van zwarte onderzoekers in de industrie, is er te weinig samenwerking tussen verschillende academische onderzoeksgebieden. AI-onderzoek zou nog te veel gedomineerd worden door computer scientists. Met name de visie vanuit de sociale en humanitaire wetenschappen wordt nog te weinig meegenomen. AI heeft namelijk impact op alle facetten van het leven.

Conclusies en vooruitblik

Heel veel dingen zijn te reduceren tot een formule met een succesvolle output, zelfs kunst, muziekcomposities en grappen. Het blijft echter narrow en specialistisch. En hoewel deze toepassingen steeds krachtiger en indrukwekkender worden, kun je je afvragen of dit daadwerkelijk intelligentie is. We noemen onze mobiele telefoons niet voor niets smartphones en geen *intelligent phones*. Laat staan *wise phones*. Hetzelfde geldt voor steden, we noemen ze *smart cities*. Intelligent zijn is iets anders dan slim zijn. Mensen kunnen goed functioneren in een nieuwe omgeving, ze kunnen generaliseren, hebben een intuïtief begrip van natuurwetten en causaliteit en houden moeiteloos rekening met de psychologische en sociologische context. Daarin loopt de ontwikkeling van AI nog tegen een muur en de verwachting is dat deze muur met deep learning niet omvergeworpen wordt. Het menselijk brein heeft dan ook een evolutie van miljoenen jaren moeten doorlopen om zo ver te komen. Waarom verwachten we van AI dan wonderen? Ook leren kost tijd; je kunt een kind van 1 jaar niet in een auto laten rijden. En als een kind zijn eerste woordje zegt zijn we dolblij, maar het kind heeft dan ook nog geen begrip van wat hij zegt. Leren gaat met vallen en opstaan.

Tegelijkertijd zijn de meeste begrippen die we onszelf toekennen, zoals intelligentie maar ook bewustzijn en vrije wil, niet onomstreden. We weten niet goed hoe het precies werkt en kunnen filosofisch gezien nauwelijks een onderscheid maken tussen onze drijfveren en de doelen die we machines meegeven. Toch blijft het nastreven van menselijke intelligentie met AI als stip op de horizon staan. Wanneer we menselijke *performances* verwachten van machines, dan lijkt het erop dat we ook de menselijke eigenschappen ervan moeten accepteren. Menselijke uitleg is namelijk ook twijfelachtig. Dat is geen neurologisch verhaal, waarbij je onder de motorkap kunt kijken. Het is een reconstructie achteraf.

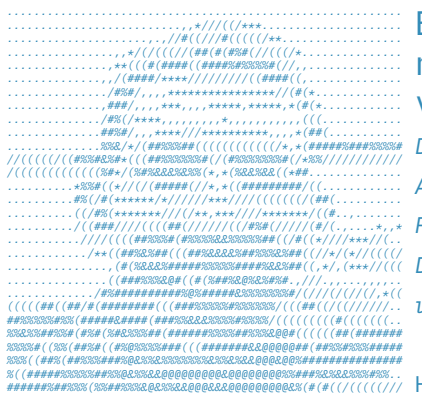
» We verwijten de computer menselijke trekjes. «

-- Patrick van der Duin, STT

Dit is overigens geen vrijbrief om de controle los te laten. Transparantie van het proces moet gewaarborgd blijven. Evenals de uitleg. Echter zullen we hoogstwaarschijnlijk wel moeten accepteren dat de uitlegbaarheid subjectief zal blijven en niet altijd volledig herleidbaar is. Daarnaast blijft het vraagstuk omtrent aansprakelijkheid een grote uitdaging.

Vanuit dat perspectief is het veel interessanter om te kijken naar de wetmatigheden van intelligentie. Bij veel technologische innovaties zijn er pas écht stappen gemaakt toen we zijn gestopt met het letterlijk proberen te kopiëren van fysiologische mechanismen. Een vliegtuig fladdert niet met zijn vleugels en een duikboot zwemt niet. Als we de wetmatigheden van intelligentie beter kunnen begrijpen zullen er grote stappen gemaakt kunnen worden. Misschien niet in het vervangen van menselijke intelligentie, maar juist in het aanvullen ervan.

De vraag blijft hoe de technologie zich in de toekomst zal gaan ontwikkelen. Wat zeggen de experts? En kan AI ons daarbij helpen? Het is hierbij van belang om niet alleen naar de technologie te kijken, maar ook naar de belangen die de ontwikkelingsrichting beïnvloeden, zoals economische en politieke belangen.



Begrip van de mens is belangrijk voor AI

Door Leendert van Maanen, Assistant professor Psychological Methods Department, Universiteit van Amsterdam

Het kraken van de geheime codeermachine Enigma tijdens de Tweede Wereldoorlog heeft ons veel geleerd over hoe mensen beslissingen nemen, maar laat ons ook zien waar computers en mensen (nog) sterk van elkaar verschillen. Het Duitse leger gebruikte de Enigma's om berichten versleuteld te verzenden. Voor het beëindigen van de oorlog was het breken van die code dus van groot belang. Om die reden werd in 1942 de Britse wiskundige Alan Turing gevraagd zich over het breken van de code te buigen. Misschien wel zijn belangrijkste bijdrage hieraan was de ontwikkeling van een algoritme om te beslissen of twee onderschepte berichten van dezelfde Enigmamachine afkomstig waren. Het algoritme scande letter voor letter twee gecodeerde berichten en bepaalde voor elk letterpaar wat de kans was dat die door toeval op dezelfde plaats in de twee berichten stonden. Teveel van die toevalligheden was een aanwijzing dat de berichten door dezelfde Enigmamachine verstuurd waren en het algoritme telde al die kansen bij elkaar op om zo bewijs te verzamelen voor een van twee uitkomsten: óf de berichten kwamen van dezelfde machine (en in dat geval werden ze geselecteerd voor verdere handmatige decodering), óf ze kwamen van verschillende machines (en in dat geval werden ze verworpen).

> Opvallend genoeg lijkt ditzelfde mechanisme door mensen gebruikt te worden om keuzes te maken. In een inmiddels klassiek artikel uit 2002 laten Joshua Gold en Michael Shadlen zien hoe de hersenen op dezelfde wijze informatie coderen als Turings algoritme dat bijna 80 jaar geleden deed. Het brein verzamelt over tijd bewijs voor de verschillende keuze-opties totdat er voldoende evidentie is om zich te committeren aan de meest waarschijnlijke optie. Het wiskundige principe dat ten grondslag ligt aan Turings algoritme wordt daarom vaak gebruikt om menselijk keuzegedrag te onderzoeken.

Wat opvalt als je keuzegedrag van mensen onderzoekt, is dat kiezen tijd kost en dat keuzes nooit exact hetzelfde zijn. Zelfs als de omstandigheden identiek zijn, kan het keuzeproces zich anders over de tijd ontvouwen. Een belangrijke oorzaak hiervan is dat sommige factoren die een keuze beïnvloeden onbekend zijn. Aandachtsprocessen, geheugenprocessen, emoties, sentimenten en context spelen allemaal een rol, waardoor het onmogelijk is precies te voorspellen hoe de keuze zich ontwikkelt. De brokjes bewijs worden telkens op subtiel andere wijze gewogen, waardoor het keuzeproces steeds anders is. Dit lijkt fundamenteel te verschillen van AI-gestuurde keuzes, waarbij de duur van een keuzeproces geen rol speelt en de uitkomst eigenlijk vaststaat. De enige belemmering voor een onmiddellijke keuze is de snelheid waarmee informatie beschikbaar komt. Maar zolang die informatie (exact) hetzelfde is, zal ook de keuze gelijk blijven: computers hebben geen spanningsboog, stemmings-wisselingen of sentimenten die van invloed zijn op hun gedrag. Als in de toekomst AI-systemen en mensen samen gaan werken, wordt het belangrijk om rekening te houden met de duur en variabiliteit van menselijke keuzes: de machine moet geduldig zijn met de grillige mens om goed te kunnen inschatten wat zij wil. Als we omgekeerd willen dat de mens de keuzes van de machine accepteert, dan is het noodzakelijk om – waar mogelijk – menselijke variabiliteit in AI in te bouwen, om op die manier de herkenbaarheid en acceptatie van de keuzes te vergroten.

> Voor een betere interactie tussen mens en machine en voor toekomstige ontwikkelingen binnen AI is onderzoek naar mensen daarom paradoxaal genoeg van cruciaal belang. Precies omgekeerd aan hoe ooit het kraken van de Enigmacode bijdroeg aan meer inzicht in menselijke keuzes.

3 Hoe gaat AI zich in de toekomst ontwikkelen?

Pagina's: 41
Woorden: 9.148
Leestijd: ca. 1 uur

Hoe gaat AI zich in de toekomst ontwikkelen?

Op basis van 40 expertinterviews en een uitgebreide literatuurstudie is een ding wel duidelijk geworden; de voorspellingen over de wijze waarop AI ↴ zich in de toekomst gaat ontwikkelen zijn enorm divers. Technologische doorbraken laten zich historisch gezien dan ook slecht voorspellen. In het boek *Profiles of the Future* beschrijft Arthur C. Clarke in 2000, naast zijn eigen toekomstvoorspellingen, voorspellingen van anderen uit het verleden. Denk aan voorspellingen ten aanzien van ruimtevaart en kernenergie. Het patroon hierbij is dat de technologie steeds door slechts een kleine groep optimisten werd voorspeld en dat een zeer grote groep met gevestigde experts de voorspellingen tegenspraken. Hierdoor kwamen de meeste technologische doorbraken uit de lucht vallen, ook voor de experts. De vraag is of dit patroon ook geldt voor AI.

De ontwikkeling van de technologie ligt echter niet alleen in handen van wetenschappers en techneuten. Er zijn economische en politieke belangen die een grote rol spelen in de ontwikkelingsrichting van de technologie. Volgens sommigen gaat Europa de boot missen en de wedstrijd verliezen van China en Amerika. Maar is het wel een wedstrijd? Of zijn er andere redenen om te investeren in de technologie? Vanuit dat perspectief is het tevens interessant om te kijken naar de positie van Nederland in het hele veld. Wat is onze strategie?

Expert-voorspellingen

Er wordt druk gespeculeerd over de toekomst van AI. Wanneer gaan we general AI ↴ bereiken? En hoe groot is de stap naar superintelligence ↴ dan? Over het feit dát we general AI gaan bereiken lijken de meeste experts het wel met elkaar eens te zijn. Wetenschappers zien hiervoor met name geen filosofische belemmeringen en benadrukken dat ons brein en een computerbrein in dat opzicht niet zoveel van elkaar verschillen.

> *» Het brein is net als een computer een fysisch mechanisme. Tenzij je denkt dat er iets ontastbaars in ons is, een ziel bijvoorbeeld. Dat mag, maar als wetenschapper zie ik dat niet zo. «*
>
> -- Marcel van Gerven, Radboud Universiteit

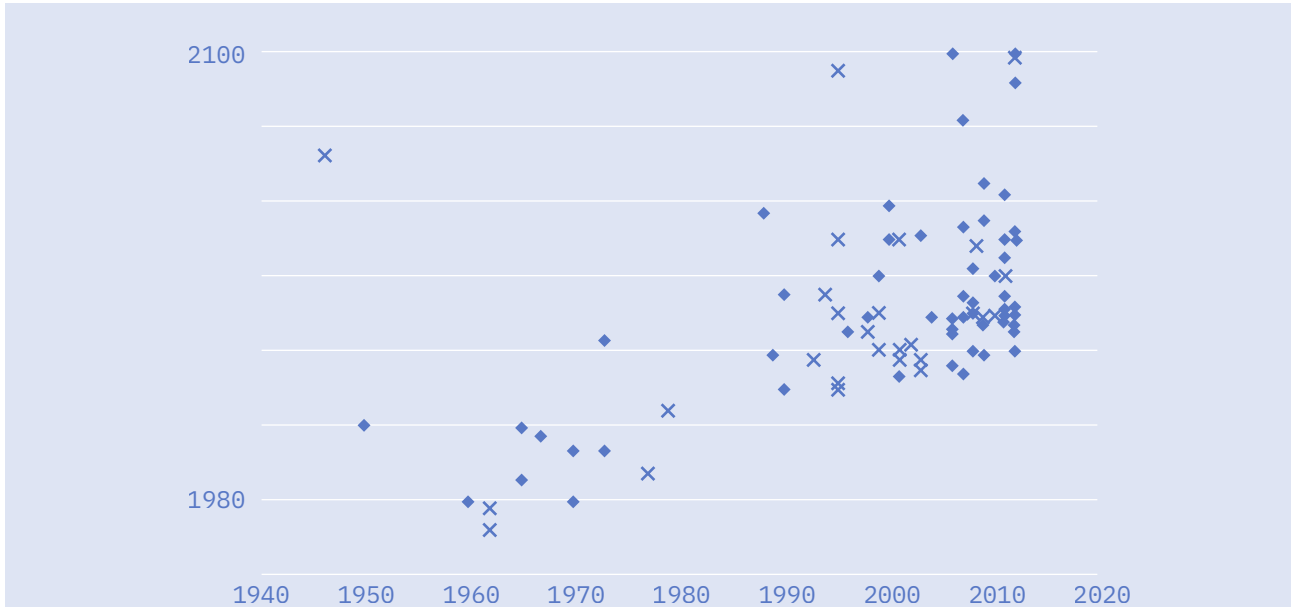
Mensen hebben de stip nu eenmaal op de horizon gezet en met het huidige momentum en de energie die daarachter zit, zou het best wel eens kunnen gaan lukken om general AI te gaan bereiken. De snelheid van de ontwikkelingen op het gebied van AI in de afgelopen decennia is ongekend. In 2005 schreef futurist Ray Kurzweil het boek *The singularity is near*. Kurzweil beschrijft hierin dat het punt waarop AI meer invloed krijgt op de ontwikkelingsrichting van de samenleving dan de mens zelf relatief dichtbij is. Als dat punt bereikt is zal AI krachtiger zijn dan alle menselijke intelligentie gecombineerd. Een belangrijke driver voor deze zogenaamde *AI Takeoff* is het potentiële vermogen van AI om zichzelf te verbeteren door zijn eigen broncodes te herschrijven en optimaliseren zónder tussenkomst van de mens. Aanvankelijk zal het intelligentieniveau van het systeem minimaal zijn, maar door continue iteraties zal het uiteindelijk evolueren tot human level AI ↴ of deze zelfs overstijgen. Er wordt hierbij onderscheid gemaakt tussen twee verschillende

takeoffs. De zogenaamde *Soft Takeoff* staat voor een takeoff waarbij het jaren of decennia kost om human-level AI te bereiken. Deze 'trage' ontwikkeling kan bijvoorbeeld het gevolg zijn van de beperking dat AI afhankelijk is van feedback uit de fysieke wereld die alleen in real time verwerkt kan worden. Experts zoals AI-wetenschapper Hans Moravec en futurist Ray Kurzweil geloven dat deze takeoff de voorkeur moet hebben, omdat deze veilig is en goed beheersbaar blijft voor de mens. Aan de andere kant bestaat ook de mogelijkheid voor een *Hard Takeoff*. Bij deze benadering zou general AI in maanden, dagen of zelfs minuten bereikt kunnen worden. Er is hierbij sprake van een *intelligence explosion*, waarbij een systeem zijn eigen proces dat intelligentie voortbrengt kan analyseren en verbeteren. Hierdoor ontstaat een intelligenter systeem dat het proces herhaalt, enzovoort. Uiteindelijk wordt er dan een 'limiet' bereikt, die volgens deze benadering menselijke intelligentie ver zal overstijgen. Deze visie wordt gedeeld door experts als futurist Michael Anissimov en filosoof Nick Bostrom.

> *» Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion," and the intelligence of man would be left far behind. «*
>
> -- I.J. Good, 1965

De verschillende opvattingen van experts worden nog duidelijker wanneer er gevraagd wordt om een concreet jaartal aan de voorspelling te koppelen. Deze voorspellingen worden ook wel *timeline predictions* genoemd. De voorspellingen lopen uiteen van 'binnen 10 jaar', tot 'over 100 jaar' en alles wat daartussen zit. Uit onderzoek van de University of Oxford uit 2014 blijkt dat deze diversiteit in opvattingen van alle tijden is. Uit 257 voorspellingen over de toekomst van AI in de

periode tussen 1950 en 2012, zijn 96 timeline predictions gefilterd en geplot in een grafiek. Hieruit blijkt dat er weliswaar geen directe correlatie is tussen de verschillende voorspellingen, maar dat het moment waarop we general AI zouden moeten gaan bereiken wel steeds vooruitgeschoven wordt. Meer dan een derde van de voorspellingen gaat ervanuit dat general AI bereikt zal worden in een periode van 15-25 jaar vanaf het moment dat de voorspelling is gedaan. Daarbij lijken experts over het algemeen iets voorzichter te zijn in hun voorspellingen dan non-experts. Hoe meer je weet, hoe meer besef je krijgt van wat je allemaal niet weet.



'Median estimate for human-level AI, graphed against date of prediction'
– Bron: University of Oxford (2014)

Futurist Martin Ford stelde in 2018 een lijst samen van 18 wereldwijd gerespecteerde experts op het gebied van AI en interviewde ze voor zijn boek *Architects of Intelligence*. Als onderdeel van deze gesprekken stelde hij de vraag wanneer de experts verwachten dat general AI bereikt zal worden. De voorspellingen lopen uiteen van 2029 (11 jaar vanaf 2018) tot 2200 (182 jaar vanaf 2018). Het gemiddelde ligt op 81 jaar. Verschillende experts gaven aan liever geen voorspelling te doen, omdat

het pad naar human-level AI zeer onzeker is en het nog onduidelijk is welke hordes er precies genomen moeten worden. De meeste experts wilden alleen anoniem een voorspelling geven. Opvallend is dat juist de experts met de twee uitersten een voorspelling *on the record* wilden geven, namelijk futurist Ray Kurzweil (2029) en MIT professor Rodney Brooks (2200). Het gemiddelde van 81 jaar is in verhouding tot andere onderzoeken overigens vrij pessimistisch. Daarbij komen de meeste clusters uit op 20-40 jaar. Bij deze onderzoeken ligt het aantal deelnemers echter wel een stuk hoger en zijn naar alle waarschijnlijkheid ook experts buiten het vakgebied geraadpleegd.

» If you want to see a true thinking machine, eat your vegetables. «

-- Martin Ford, 2018

De grootste misvattingen

Met name toekomstvisies die gepaard gaan met de harde takeoff hebben de neiging om ongewenste scenario's op te roepen. Denk bijvoorbeeld aan machines die alle banen overnemen of de wereld willen besturen. Volgens Rodney Brooks leiden deze 'verkeerde voorspellingen' tot angsten voor dingen die 'nooit gaan gebeuren'. Volgens hem zijn er zeven redenen waarom deze toekomstvoorspellingen zo uit de bocht vliegen:

1: We overschatten de effecten van de technologie op de korte termijn en onderschatten de gevolgen ervan op de lange termijn.
2: We doen voorspellingen die zo ver af staan van de mogelijkheden van de huidige technologie, dat we de beperkingen ervan niet kunnen overzien. Hierdoor zijn dergelijke uitspraken niet goed te toetsen.
3: We vergelijken de specialistische capaciteiten van machines met algemene competenties van mensen. We gaan er hierbij ten onrechte vanuit dat machines hun capaciteiten ook kunnen toepassen op andere taken.
4: We vergeten dat machine learning ↴ veel menselijke voorbereiding vereist. Het is gebaseerd op specifieke codes en datasets die ontworpen zijn voor een specifiek doel binnen een specifieke context. Deze kunnen daardoor niet zomaar in andere domeinen worden toegepast.
5: We verwachten dat de ontwikkelingssnelheid van technologie zich continu ontwikkelt en negeren het feit dat het vaak om geïsoleerde doorbraken gaat die moeilijk te voorspellen zijn.
6: We gaan ervanuit dat de wereld – op een specifieke en plotselinge twist na – in de toekomst in essentie exact hetzelfde is als vandaag. Wanneer we echter in staat zijn om superintelligente machines te ontwikkelen, dan zal de wereld als gevolg van een geleidelijke evolutie significant anders zijn.
7: We verwarren de snelle integratie van nieuwe software met die van nieuwe hardware. In de praktijk duurt het echter veel langer voordat oude systemen volledig vervangen zijn en in de breedte van de samenleving worden gebruikt.

Met name punt 6 wordt vaak over het hoofd gezien. Dit zie je bijvoorbeeld terug in de populaire Netflix-serie *Black Mirror*. De ontwikkeling van één specifieke technologie wordt enorm vooruit gespoeld, terwijl de rest van de omgeving gelijk blijft. Dit verklaart enerzijds het succes van de serie, mensen kunnen zich er op deze wijze makkelijker in verplaatsen, maar het geeft anderzijds een onrealistisch beeld. De context zou namelijk mee moeten veranderen. Zo werd de zelfrijdende auto in de jaren '50 al voor het eerst gevisualiseerd. Men dacht destijds dat dit meer tijd en ruimte zou bieden voor gezinnen om samen gezelschapspellen te spelen. Huidige visualisaties van de zelfrijdende auto gaan veel meer in op efficiënte en tijdsbesparing, zodat we onderweg kunnen werken; *work on wheels*. De focus op de technologie zorgt ervoor dat voorspellingen vaak niet uitkomen, doordat de veranderende context onvoldoende meegenomen wordt.

If you want to see a true thinking machine, eat your vegetables.

-- Martin Ford, 2018

» AI will not replace human labor, instead it will change the nature of jobs and create many new opportunities that were unimaginable before. «

-- Max Welling, UvA

Uitdagingen voor de toekomst

Of het nog 10 jaar gaat duren of nog bijna 200 jaar, met de huidige technologieën benaderen we menselijke intelligentie hoogstwaarschijnlijk niet. Hoe relevant en krachtig ze ook zijn, de huidige toepassingen van AI zijn specialistisch en narrow ↴. Een van de grootste technologische uitdagingen volgens experts is *transfer learning*. Een machine kan leerervaringen ↴ binnen een specifiek domein nog niet toepassen op een gebied in een ander domein. Zo kan Google Translate verkeersregels prima vertalen, maar het heeft geen idee hoe ze toegepast moeten worden in het verkeer. De flexibiliteit en common sense ↴ die mensen hiervoor nodig hebben is voor machines nog een enorme uitdaging. Geïntegreerde informatie gaat verder dan statistiek. Bijvoorbeeld bij medische diagnoses: standaard beslissingen zijn zo over te nemen, maar bij meer zeldzame ziektes wordt het lastig. Daar is te weinig data voor. Een computer kan wel kankercellen herkennen op basis van afbeeldingen en dit veel sneller en efficiënter doen dan het menselijk oog. Maar een machine zou geen complexe operaties kunnen uitvoeren, waarbij er verbanden moeten worden gelegd die niet uit patronen bestaan (Transformaties ↴).

Een andere grote uitdaging is causaliteit. Data kan soms onverklaarbaar zijn. Bijvoorbeeld bij een verzekeraar: data wijst uit dat rode auto's gemiddeld hogere verzekeringsclaims hebben. Maar komt dit omdat mensen met rode auto's agressiever rijden, of zit er een rode Lamborghini in de dataset die *total loss* is gereden en het gemiddelde scheef trekt? Zonder menselijke tussenkomst zou de premie voor mensen met een rode auto zomaar automatisch verhoogd kunnen worden. AI is niet goed in het verschil

tussen correlatie en causaliteit. Het kan weliswaar samenhang tussen twee gebeurtenissen zien, maar het kan niet goed vaststellen of de ene gebeurtenis ook daadwerkelijk het gevolg is van de gebeurtenis die daaraan vooraf is gegaan. Wanneer er in de zomer meer ijsjes worden verkocht en er tegelijkertijd ook meer verdrinkingen zijn, dan vertelt ons gezond verstand dat hier geen oorzakelijk verband tussen zit. Voor een computer is dit veel ingewikkelder.

Daarnaast vormt de beperkte Kennis over ons brein ↴ een uitdaging voor het bereiken van general AI. De wijze waarop de hiërarchie van taken en geheugen in ons brein bij leren precies werkt weten we niet. Naast de fysiologische en mechanische structuur is ons brein ook nog omgeven door chemische processen en hebben emoties invloed op ons leerproces. We hebben hoop, verwachtingen en ambities. Daarnaast zijn we sociale wezens; we werken samen en leren van elkaar in verschillende contexten. Dergelijke leerervaringen zijn niet zomaar 'na te maken'.

» Het gapende gat middenin AI is de definitie van I. Dat is pijnlijk. «

-- Frank van Harmelen, VU

De uitdagingen voor de toekomst worden door het onderzoek van psychologen Felix Warneken en Michael Tomasello op een treffende manier duidelijk. Ze deden een reeks van experimenten om te onderzoeken op welke wijze altruïsme werkt bij jonge kinderen. Het experiment zag er als volgt uit:

Een man loopt met een stapel boeken naar een kast met gesloten deuren. Bij de kast aangekomen loopt de man met zijn boeken tegen de kastdeuren aan. Een peuter kijkt vanuit de hoek van de kamer toe en ziet hoe de man terugloopt om vervolgens opnieuw met de boeken in zijn handen tegen de gesloten kastdeuren aan te lopen. Na twee pogingen grijpt het jongetje in en opent de kastdeuren. Nadat hij de deuren geopend heeft kijkt hij naar de man en

vervolgens naar de boekenplank. De man legt de boeken vervolgens in de kast.

In de gehele scene wordt niet gepraat en toch is het jongetje heel bewust van de bedoeling van de man. Allereerst heeft dit te maken met bewustzijn ↴ en empathie ↴; het jongetje kan zich in de man verplaatsen en kan daaruit het doel van de man afleiden. Dat lijkt niet zo heel bijzonder, maar dat is het wel. Het doel is op zichzelf namelijk niet expliciet en is niet duidelijk waarneembaar. In theorie zou de man ook hele andere doelen kunnen hebben. Hij zou ook kunnen proberen om de stapel boeken op de kast te leggen of met behulp van de boeken de kast om te duwen. Het feit dat het jongetje hier niet vanuit gaat heeft te maken met common sense. Zo weet het kind bijvoorbeeld dat een kast hol is en niet massief. En dat het planken heeft waar je boeken op kunt leggen. Dit gaat dus niet over fysiek waarneembare aspecten, maar over abstracte concepten. Hierbij speelt tevens het zogenaamde *frame problem*: het jongetje bepaalt intuïtief welke informatie relevant is en welke informatie niet. Zo is de kleur of de vorm van de kast voor het jongetje niet relevant, maar wel het feit dat het twee deuren heeft die geopend kunnen worden. Ook al heeft het jongetje deze letterlijke situatie hoogstwaarschijnlijk nog nooit in zijn leven meegemaakt, zijn gezond verstand stelt hem in staat om op de juiste manier te handelen.

Wanneer je een computer wilt programmeren om vergelijkbare taken te kunnen doen, dan loop je tegen een rekenkundig probleem aan. Het aantal mogelijkheden is namelijk bijna oneindig, waardoor het zelfs met de krachtigste computers nog niet exact is op te lossen. Dit wordt ook wel computationele complexiteit genoemd. Daarbij kunnen computers door middel van deep learning ↴ objecten weliswaar herkennen, maar ze weten niet precies wat het object daadwerkelijk betekent. Computers hebben geen intentionaliteit ↴.

» We're worried about something taking over the world that can't even on its own decide to play chess again. «

-- Oren Etzioni, Allen Institute for Artificial Intelligence

Doorbraken in de toekomst

De meeste ontwikkelingen op het gebied van AI zijn nu nog te danken aan enorme datasets waarmee machines getraind worden. Maar mensen leren niet op deze wijze. Wanneer we iets zien, beschrijven we het met behulp van taal. Wanneer we iemand ergens over horen praten, stellen we ons voor hoe deze objecten en gebeurtenissen er in de echte wereld uitzien. Mensen leven in een fysieke wereld, vol met sensorische prikkels. Deze ervaringen maken het voor mensen mogelijk om de wereld te begrijpen. Dit principe wordt ook wel symbol grounding ↴ genoemd.

» We don't leave our babies with an encyclopedia in the crib, expecting them to understand the world. «

-- Boris Katz, MIT

Het huidige AI-onderzoek is in de afgelopen decennia uiteengevallen in twee verschillende kampen. De 'symbolisten' die intelligente machines proberen te bouwen door ze te coderen met formele logica ↴ en symbolische informatie en de 'connectionisten' die met behulp van op biologie geïnspireerde neurale netwerken ↴ machines proberen te trainen met grote hoeveelheden data. Traditioneel gaan deze twee kampen niet zo goed samen. Maar ironisch genoeg ligt de verwachte doorbraak binnen AI juist in de combinatie van deze twee benaderingen, namelijk *Deep Reasoning*. Neurale netwerken zorgen ervoor dat het systeem op basis van patroonherkenning kan 'zien' en de symbolische programmering zorgt ervoor dat het systeem kan 'redeneren'. Deze vorm van leren lijkt op de wijze waarop kinderen leren, namelijk door te kijken én te praten.

> *» Instead of trying to produce a programme to stimulate the adult mind, why not rather try to produce one which simulates the child's? «*
>
> -- Alan Turing, 1950

Onderzoek van MIT, IBM en DeepMind uit 2019 toont de potentie van deze benadering. Zij ontwikkelden de *Neuro Symbolic Concept Learner* (NS-CL), een systeem dat uit verschillende AI-technieken bestaat. Een neuraal netwerk is getraind op een reeks situaties met verschillende objecten. Een ander neuraal netwerk is getraind op vraag-antwoord gesprekken over deze situaties. Op deze wijze kan het vragen beantwoorden over de situaties. Het systeem is daarbij geprogrammeerd om symbolische concepten in de teksten te begrijpen, zoals objecteigenschappen en ruimtelijke relaties. Deze kennis helpt het systeem om nieuwe vragen te beantwoorden over andere situaties. Het systeem kan dus concepten herkennen in nieuwe vragen en deze koppelen aan visuele objecten uit eerdere situaties. Zónder hiervoor opnieuw getraind te worden. Met de huidige neurale netwerken is dit niet mogelijk.

> Deze nieuwe benadering pakt problemen van de oude benaderingen aan door ze te combineren. Deep Reasoning verhelpt het schaalbaarheidsprobleem van het symbolisme (het is onmogelijk om alle opties efficiënt te programmeren) en het pakt tegelijkertijd het dataprobleem van neurale netwerken aan (grote datasets zijn vaak niet beschikbaar of onvolledig). Op deze wijze kunnen bijvoorbeeld robotsystemen in real time leren, zonder dat ze veel tijd kwijt zijn met trainen in specifieke omgevingen.

Onderzoek in de toekomst

Ontwikkelingen op het gebied van AI brengen machines met een vorm van 'agentschap' steeds dichterbij. Oftewel, machines die zelfstandig acties ondernemen en autonome besluiten nemen. Machines vormen hiermee een nieuwe klasse van actoren in onze samenleving met hun eigen gedragingen en ecosystemen. Dit vraagt volgens experts om de ontwikkeling van een nieuw wetenschappelijk onderzoeksgebied, namelijk die van *Machine Behavior*. Onderzoekers van verschillende Amerikaanse universiteiten en bedrijven hebben in 2019 een artikel gepubliceerd in *Nature* waarin ze oproepen tot een interdisciplinaire wetenschappelijke onderzoeksagenda. Het doel is om meer inzicht te krijgen in het gedrag van artificiële intelligente systemen. Momenteel concurreren wetenschappers van verschillende disciplines nog te veel met elkaar. AI-experts en system engineers moeten meer gaan samenwerken met neurologen, biologen, psychologen, filosofen, sociologen, economen etc. Het uitgangspunt hierbij is dat we AI-systemen op dezelfde manier moeten bestuderen als de manier waarop we dieren en mensen bestuderen, namelijk door middel van empirische observatie en experimenten. Bij mensen en dieren kunnen we namelijk ook niet onder de Motorkap ↴ kijken en verwachten dat alle processen te volgen zijn.

> *» Maybe we don't need to look inside the black box after all. We just need to watch how machines behave, instead. «*
>
> -- Karen Hao, MIT Technology Review

Machines als zelfstandige agents klinkt misschien als toekomstmuziek, maar het is nu al gaande. We hebben steeds meer interactie met zelfdenkende machines. Machines die informatie vergaren, hiervan leren, dit delen met andere apparaten in het systeem en hierdoor steeds complexere besluiten kunnen nemen. Denk bijvoorbeeld aan Alexa, de home assistent van Amazon. Het apparaat leert op basis van gesprekken en kan hierdoor specifieke aanbevelingen geven die het gedrag van gebruikers beïnvloedt. Het is steeds moeilijker om te herleiden of deze aanbeveling geoptimaliseerd is op commerciële doelen, of bedoeld is om mensen van het best passende advies te voorzien.

Economische en politieke belangen

De ontwikkeling van technologie is voor een groot deel afhankelijk van investeringen. De AI-winters ↴ in het verleden werden dan ook voornamelijk veroorzaakt door een gebrek aan financiering voor AI-onderzoek. Veel van deze financieringen komen vanuit de overheid. Momenteel is de interesse in AI ↴ dermate groot dat wereldmachten de strijd met elkaar aangaan. Uit onderzoek van PwC uit 2017 blijkt dat het wereldwijde *Gross Domestic Product* (GDP) in 2030 door ontwikkelingen op het gebied van AI maar liefst 14% hoger zal liggen.

> *» Whoever leads in AI will rule the world. «*
>
> -- Vladimir Poetin, 2017

De Russische president Poetin ziet enorme kansen voor AI in de toekomst, maar waarschuwt ook voor dreigingen die moeilijk te voorspellen zijn. Volgens Poetin zal het land met de beste AI over de wereld heersen. We moeten daarom oppassen voor het ontstaan van een monopolie. De race lijkt voornamelijk tussen de VS en China te gaan. Maar welke rol neemt Europa hierin? En hoe verhoudt de Nederlandse strategie zich tot andere landen?

De vraag daarbij is in hoeverre de overheid nog grip heeft op de ontwikkeling van de technologie. Bedrijven als Google, Amazon en Facebook worden steeds machtiger en krijgen steeds meer invloed.

> *» History belongs to the highest bidder now. «*
>
> -- Viktor Petrov, House of Cards

Laat de leeuw niet in zijn hempie staan

In 2018 verscheen het rapport AI voor Nederland. Een eerste aanzet voor een Nederlandse AI-strategie. Het rapport is opgesteld door AINED, een publiek-private coalitie van TopTeam ICT, VNO-NCW, ICAI, NWO en TNO.

Het uitgangspunt van het rapport is dat de mogelijkheden die AI nu en in de toekomst kan bieden in Nederland nog onvoldoende benut worden. AI zou een nationale prioriteit moeten worden, om zo onze welvaart en internationale concurrentiepositie voor de toekomst veilig te kunnen stellen. Nederland moet zich dus gaan positioneren en actie ondernemen. Het voorstel van AINED voor een nationale strategie is als volgt:

1. Verhoog de capaciteit van het (AI-)onderwijs
2. Trek buitenlands talent aan
3. Ontwikkel AI-kennis bij executives en beroepsbevolking
4. Verhoog bruikbaarheid, beschikbaarheid en waarde van publieke data
5. Ontwikkel (op lange termijn) een PPS platform voor data delen
6. Verbeter de toegang tot WBSO-subsidie voor AI innoverende bedrijven
7. Faciliteer de toegang tot Europese investeringen en netwerken
8. Zet een groeifonds voor AI op bij Invest-NL
9. Zet in op ruimere toegankelijkheid tot financiering en verhoogde awareness omtrent AI binnen het MKB
10. Pas AI toe binnen de eigen processen van de overheid
11. Ontwikkel gericht de AI-competenties binnen de overheid
12. Bepaal kaders in samenwerking met Europa en werk kaders uit op sector- en waar nodig casusniveau
13. Richt publiek-private samenwerking en financiering in voor data- en kennisgeoriënteerde waardeketens
14. Zet een onderzoekscentrum op gericht op high-tech onderzoek samen met kennisinstellingen en organisaties via industriële kennislaboratoria

In het rapport wordt benadrukt dat Nederland dit niet alleen kan. Nederland moet aansluiten – en waar mogelijk leiderschap tonen – bij Europees beleid, programmering en allianties. Het rapport onderstreept dat voor een succesvol Europees beleid er een nationale strategie

WHO
EVER
LEADS
IN AI WILL
RULE THE WORLD:

-- Vladimir Poetin, 2017

nodig is die alle aspecten zichtbaar verenigt en keuzes maakt voor Nederland. Hiervoor is politieke steun nodig om op het moment van besluitvorming in Europa de Nederlandse belangen te kunnen bestendigen.

Het standpunt van de overheid

Om een goed beeld te krijgen van de mogelijke impact van AI op de publieke waarden heeft Mona Keijzer, Staatssecretaris van Economische Zaken en Klimaat, in 2018 namens (bijna) alle ministeries een adviesaanvraag ingediend bij de Wetenschappelijke Raad voor Regeringsbeleid (WRR) om nader onderzoek te doen. Het kabinet vindt het van belang te bezien welk instrumentarium reeds voorhanden is om de kansen van Al te faciliteren en de uitdagingen te beantwoorden, en welke instrumenten eventueel aanvullend nodig worden geacht. De Nederlandse overheid komt voor de zomer van 2019 met een Strategisch Actieplan AI (SAPAI). Met name op het gebied van Open Data voorziet Nederland een sterke positie voor zichzelf weggelegd.

Toch is de Nederlandse overheid in relatie tot andere Europese landen vrij laat met een nationale AI-strategie. Finland publiceerde al in 2017 Finland's Age of Artificial Intelligence. In 2018 volgenden ook Denemarken, Italië, Frankrijk, Verenigd Koninkrijk, Zweden en Duitsland. Net voor het eind van dat jaar kwam ook de Europese Commissie met een gecoördineerd plan om een boost te geven aan AI 'made in Europe'. Dit lijkt een gemiste kans voor Nederland om leiderschap te tonen bij de totstandkoming van Europees beleid, programmering en allianties.

Koude oorlog 2.0

Het zwaartepunt van de wereldeconomie verschuift naar het Oosten. Het principe van 'waar wonen de meeste mensen' gaat weer tellen. China had in 2017 al twee strategische AI-plannen klaar, namelijk het Three-year action plan en het Next Generation AI plan. In dit laatste plan beschrijft China haar doel om in 2030

wereldleider te zijn in AI. Uiteraard wil de VS zich niet van de troon laten stoten en kwam in de eerste helft van 2018 met het White House Summit on AI. De VS zet alles op alles om de voorsprong op China te behouden. Dit uit zich in The American AI Initiative.

> *» Continued American leadership in Artificial Intelligence is of paramount importance to maintaining the economic and national security of the United States. «*
>
> -- Donald J. Trump, 2019

De rivaliteit tussen de twee landen is goed zichtbaar in het gevecht over de toekomst van 5G netwerken, waar Europa middenin zit. Deze *fifth generation networks* maken een supersnelle verbinding mogelijk en zorgen ervoor dat technologische toepassingen veel beter functioneren, van mobiele telefoons tot zelfrijdende auto's. Dit zal een grote rol spelen bij de ontwikkeling van AI-toepassingen. Het gevecht tussen de VS en China gaat met name over de vraag of Europese landen toestemming moeten geven aan de Chinese telecomgigant Huawei om 5G netwerken op het continent te installeren. Europa staat hier in de basis wel voor open aangezien de kwaliteit hoog is en de kosten laag. De VS waarschuwt er echter voor dat wanneer Europa de technologie toestaat het grote gevolgen zal hebben voor de nationale veiligheid. China zou de netwerken namelijk gebruiken voor onder andere spionage. De Amerikaanse overheid heeft daarom al aangegeven geen zaken te doen met Huawei.

Interessant hierbij is dat deze twee wereldmachten een compleet andere visie hebben op welvaart. Daar waar de VS gelooft dat welvaart groeit door bedrijven de maximale vrijheid te geven, gelooft China dat juist de overheid de maximale vrijheid moet krijgen om de welvaart te laten groeien. Hoewel deze visies lijnrecht tegenover elkaar staan hebben ze wel een overeenkomst; ze kenmerken zich beide door een proactieve en doelgedreven benadering.

Sorry, president Trump. Maar op AI-gebied gaat China winnen

door Carl Rohde, International Future Researcher

De strijd tussen de VS en China om AI-werelddominantie is volop losgebarsten en zal de komende jaren steeds opnieuw de kop opsteken. Maar wie heeft de beste papieren om de competitie te winnen? Om twee redenen zijn dat volgens International Future Researcher Carl Rohde de komende tien jaar de Chinezen.

De eerste reden heeft te maken met het verschil tussen initiëren en implementeren. AI betekent een mega-revolutie, zowel in ons maatschappelijk bestaan als in de zakenwereld. Om zo'n revolutie in eerste instantie aan te wakkeren zijn er genieën van wereldformaat nodig. Waar het AI betreft worden die vrijwel uitsluitend door Silicon Valley voortgebracht, met zijn California-cultuur van zonnig vooruitgangsoptimisme en the *sky is the limit* openheid. Steve Jobs, Elon Musk en Peter Thiel imponeren niet voor niets wereldwijd. Nu evolueren we echter naar de implementeringsfase van AI. Hoe gaat AI ingezet worden in de auto-industrie en de gezondheidsindustrie, in educatie en hospitality – en alle velden daartussen? Daarvoor zijn niet zozeer nog meer *out of the box* bevlogen genieën nodig. Maar gedisciplineerde legers van bekwame implementeer-engineers. Daarin lijkt China superieur, zowel kwalitatief als kwantitatief.

Er is een tweede reden waarom China voor nu de beste papieren heeft. AI werkt strakker en scherper naarmate het gevoed wordt met meer data. China heeft verreweg de meeste data. Zo betalen er meer mensen met hun mobiele telefoons – de ultieme dataleverancier – dan in Europa en Amerika samen. Zoals ook Weibo groter is dan Google. En Didi groter dan Uber. Daarnaast bestaat de collectieve privacy-gevoeligheid van het Westen in China niet of nauwelijks. En waar ze bestaat is ze tandeloos.

Kortom, China heeft het tij mee. Totdat er weer een nieuwe totaal disruptieve revolutie 'gemaakt' wordt. Die kan zich uitkristalliseren rondom *blockchain*-technologieën of rondom *quantum computing*. Maar nu eerst AI implementeren!

Het antwoord van Europa

In Europa ligt dat wat anders. Hier geloven we dat bedrijven en de overheid met elkaar moeten overleggen over hoe we de welvaart het best kunnen verhogen. Het is vanuit dat opzicht geen toeval dat de Europese AI-strategie zich sterk richt op aandacht voor ethische richtlijnen. In 2019 werden de Ethics guidelines for trustworthy AI gepubliceerd, ontwikkeld door een onafhankelijke *High-Level Expert Group on Artificial Intelligence*. Over de vraag of Europa hiermee een goede stap heeft gemaakt lopen de gedachten echter nogal uiteen.

Voorstanders van de Europese focus op ethiek benadrukken dat dit geen rem is voor Europa, maar juist nieuwe mogelijkheden schept en een driver van vooruitgang zal zijn. Het draagt namelijk bij aan het vertrouwen in de technologie en dat is goed voor de consumptie en dus de winst. Amerika en China zouden deze richtlijnen wel op moeten volgen, omdat Europa een grote consumentenmarkt is. Dit zie je al gebeuren bij de Europese *General Data Protection Regulation*, oftewel GDPR. Er bestaan in Europa verschillende collaboratieve initiatieven die zich inzetten voor een human-centered benadering, zoals de *Confederation of Laboratories for Artificial Intelligence Research* (CLAIRE), het *European Laboratory for Learning and Intelligent Systems* (ELLIS) en de Alliance for Artificial Intelligence (ALLAI). Deze laatste alliantie is opgezet door de drie Nederlandse deelnemers van de High-Level Expert Group on Artificial Intelligence.

Tegenstanders van de Europese focus op ethiek denken dat we voordat we zijn 'uitgepolderd' alweer achterlopen en de buit reeds verdeeld is door Amerika en China, die er

niet zoveel over nadenken maar handelen. Nu al zie je dat de unieke positie van Europa op het gebied van ethische AI links en rechts wordt ingehaald. Zo heb je in de VS al een gelijksoortig initiatief zoals ALLAI dat zich richt op AI to *benefit people and society*, namelijk het Partnership on AI. Volgens tegenstanders ontdek je juist tijdens het handelen wat gepast is en wat niet. Zaken als explainibility ↴ zijn volgens tegenstanders weliswaar belangrijk, maar ze benadrukken dat je dit niet vanaf de zijlijn kunt doen. Je moet hier in gecontroleerde omgevingen mee experimenteren. Dergelijke uitdagingen zouden haast onmogelijk theoretisch te benaderen zijn.

» We kunnen geen scheidsrechters opleiden, als we geen spelers in het veld hebben. «

-- Maarten de Rijke, UvA

Volgens Robin Young, een Chinese AI-ondernemer die zowel in China als de VS werkt, kan Europa nog veel leren van China. Europa moet volgens hem een duidelijke richting kiezen, het aandurven om risico's te nemen en werken aan een grotere publieke acceptatie van het gebruik van data. Daarbij zijn beleidsmakers in China geen advocaten (zoals in Europa vaak het geval is), maar engineers. Geen denkers, maar doeners. Volgens Young is het voor Europa lastig om met de huidige AI-technologieën te concurreren. Amerika en China zijn namelijk veel grotere homogene markten, wat een grotere beschikbaarheid van data betekent. Europa zou zich volgens hem daarom veel meer moeten gaan richten op 'de nieuwe fase van AI'. Een fase van nieuwe AI-technologieën die minder afhankelijk zijn van data ↴. Daar liggen volgens Young de kansen voor Europa.

Het is in de basis prima (en onoverkoombaar) dat de strijd om AI-technologieën vanuit een economisch en geopolitiek perspectief als een wedstrijd wordt gezien. We moeten er echter voor waken dat ethiek ook een wedstrijd wordt. Het is niet de vraag welk continent de

béste ethische richtlijnen heeft, het gaat er juist om op welke vlakken overeenstemming gevonden kan worden. Om in sporttermen te blijven; ethiek is een teamsport. Er is mondiale samenwerking nodig om in de toekomst zorg te dragen voor ethische AI die bijdraagt aan het welzijn van individuen en samenlevingen. Dit zal niet zonder slag of stoot gaan en het zal waarschijnlijk alleen lukken door een combinatie van theorie en praktijk. Ja, het is belangrijk om samen met een groep interdisciplinaire experts richtlijnen te formuleren, maar we moeten ze ook in de praktijk gaan toetsen en ervaringen uitwisselen.

Silicon Valley

In alle commotie omtrent de rivaliteit tussen China en Amerika zou je bijna vergeten dat het de grote tech-bedrijven zijn die de AI-wapenwedloop faciliteren. Zo heeft Google zowel met de Amerikaanse, als met de Chinese overheid contracten afgesloten. Onder de noemer Project Maven werkte Google vanaf 2017 voor het Amerikaanse Ministerie van Defensie aan zelflerende Beeldherkenningssytemen ↴ voor militaire drones. De software moet razendsnel vijandige objecten kunnen herkennen op grote hoogte. Hiermee komt de stap naar drones die zelfstandig beslissingen kunnen nemen en objecten kunnen uitschakelen steeds dichterbij. Het contract van Google met het Pentagon stuitte daardoor op veel weerstand, ook van Google-medewerkers zelf. In 2018 werd aangekondigd dat het contract in 2019 zou aflopen. In dezelfde periode werkte Google voor de Chinese overheid aan *Project Dragonfly*. Een zoekmachine die resultaten blokkeert die door de Chinese overheid als onwenselijk worden beschouwd. Ook hier stuitte Google op weerstand. Volgens Amnesty International zou de gecensureerde zoekmachine het recht op vrije meningsuiting en de privacy van miljoenen Chinezen bedreigen. Later kwamen ook de medewerkers van Google in opstand en schreven een open brief aan het management van Google. Google kondigde in 2018 aan de werkzaamheden stop te zetten, maar medewerkers betwijfelen of dit ook echt gebeurd is.

Het is niet zo vreemd dat overheden aankloppen bij Silicon Valley. Er is een enorm tekort aan AI-specialisten. Wereldwijd zouden er slechts 10.000 mensen zijn die AI écht begrijpen en de benodigde skills hebben om complexe problemen op te kunnen lossen. De grootste talenten worden met buitensporige salarissen weggelokt bij universiteiten door tech-bedrijven als Google, Facebook en Amazon. Dit wordt ook wel de *braindrain* van AI genoemd. Dit verstevigt de positie van deze tech-bedrijven enorm. Tech-bedrijven beheren steeds meer data die niet alleen commercieel gezien interessant zijn, maar ook voor overheden enorm relevant. Dit maakt dat overheden een bepaalde afhankelijkheidspositie ten opzichte van tech-bedrijven gaan krijgen. Het is daarom nog maar de vraag in hoeverre overheden in de toekomst een grote rol zullen spelen bij de ontwikkelingsrichting van AI.

Concurrentie tussen bedrijven heeft tevens invloed op de ontwikkelingsrichting van AI. Denk bijvoorbeeld aan de zelfrijdende auto. Dit systeem lijkt beter te functioneren wanneer alle voertuigen met elkaar in verbinding staan. De vraag is of de voertuigen van Google, BMW en Tesla dit gaan toelaten. Denk bijvoorbeeld aan de verschillende aansluitingen en besturingssystemen van mobiele telefoons. Devices van Apple en Samsung zijn allesbehalve *compatible*. Hierbij dient niet alleen naar AI gekeken te worden. De snelheid van veranderingen zit vaak in de combinatie van technologieën: AI + Cloud technologieën + Internet of Things (IoT), et cetera.

AI knows best

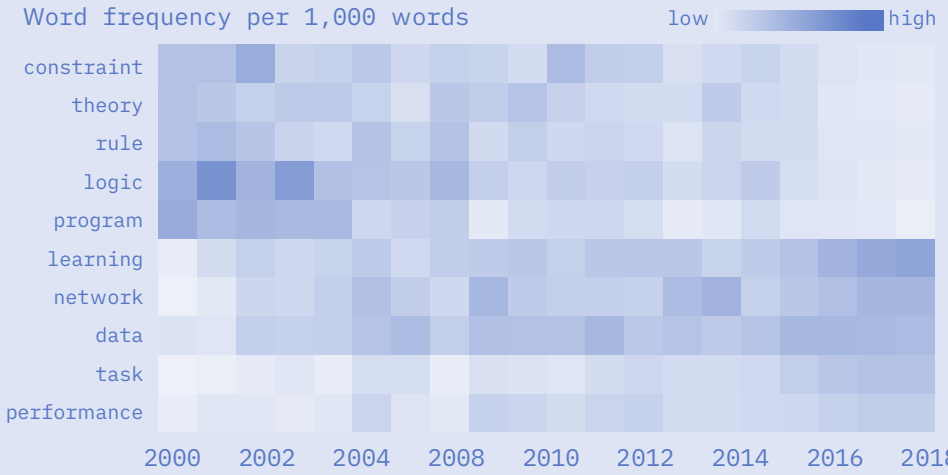
» *Wie kan de toekomst van AI beter voorspellen dan AI zelf?* «

Deze vraag stelde ik mezelf tijdens het onderzoek naar de impact van AI ↴ op besluitvorming ↴ in de toekomst. Er is enorm veel geschreven over AI en de meningen lopen sterk uiteen. Alleen al over Neurale netwerken ↴ zijn er meer dan 150.000 wetenschappelijke artikelen gepubliceerd. Het paradoxale gegeven hierbij is dat AI deze artikelen in theorie véél sneller kan analyseren dan dat ik dit als onderzoeker zelf zou kunnen. Dit inspireerde me om te gaan onderzoeken op welke wijze AI mij zou kunnen helpen bij mijn onderzoek. Ik ben daarom op zoek gegaan naar *best practices* waarbij AI is ingezet als analysetool voor onderzoek.

Wat leert de data ons?

Naast de inzet van complexe AI-technieken kan ook een relatief eenvoudige *word count* ons al veel vertellen over de ontwikkelingsrichting van AI. Met name wanneer je dit afzet tegen de tijd. Dominante termen in wetenschappelijk onderzoek geven namelijk een hint welke technieken in opkomst zijn en meer invloed zullen gaan krijgen in de toekomst. AI-onderzoek wordt gekarakteriseerd door de plotselinge opkomst en ondergang van verschillende AI-technieken. Er is hierbij vaak veel concurrentie tussen verschillende benaderingen. Zo nu en dan is er daadwerkelijk sprake van een doorbraak, waar veel wetenschappers zich achter scharen. Momenteel is dat Deep learning ↴. De vraag is echter of deze focus de komende jaren blijft of dat er een andere techniek momentum krijgt. Veel van de technieken die de afgelopen 25 jaar gebruikt zijn ontstonden in dezelfde periode, namelijk in de jaren ’50. Neurale netwerken kenden een sterke opleving in de jaren ’60 en een korte piek in de jaren ’80, maar was bijna uitgestorven totdat deep learning er in 2012 weer een nieuwe impuls aan gaf.

Het is daarom interessant om te kijken naar opkomende en stagnerende begrippen in wetenschappelijk onderzoek. MIT Technology Review selecteerde meer dan 16.000 wetenschappelijke publicaties van arXiv met het label ‘artificial intelligence’ uit de periode tussen 1993 en 2018 en analyseerde de gebruikte woorden.



‘Word frequenty per 1000 words’ – Bron: MIT Technology Review

Dit gaf enerzijds de verwachtte inzichten, namelijk de shift van kennissystemen ↴ naar machine learning ↴ eind jaren ’90 en de opkomst van neurale netwerken in 2012. Maar ook minder voor de hand liggende ontwikkelingen, zoals het stijgende momentum voor Reinforcement Learning ↴ vanaf 2015. Deze opleving is te danken aan het feit dat DeepMind’s AlphaGo in dat jaar de wereldkampioen in het eeuwenoude spel Go versloeg en getraind was aan de hand van reinforcement learning.

Wanneer je uitzoomt en kijkt naar de ontwikkelingsgeschiedenis van AI wordt er een patroon zichtbaar: elk decennium is er een andere techniek dominant. De verwachting is dat er vanaf 2020 weer een nieuwe techniek zal opkomen (of een oude techniek met een opleving). Helaas laat een analyse van begrippen niet zien welke techniek dat gaat zijn. Gebaseerd op expertvoorspellingen zal dit gaan over de combinatie van kennissystemen en deep learning. Het lijkt een kwestie van tijd voordat Deep reasoning ↴ als term zal opduiken in de literatuur.

AI onderzoekt AI

Het naar boven halen van de meest dominante begrippen in wetenschappelijke publicaties is natuurlijk zeer interessant, maar het geeft geen volledig beeld. Verschillende domeinen hanteren namelijk verschillende definities en benaderingen ten aanzien van AI. De wetenschap hanteert een ander perspectief dan de media, het onderwijs of de industrie. Daarbij is het handmatig analyseren van een grote hoeveelheid publicaties zeer tijdrovend en complex. Elsevier heeft voor haar AI-onderzoek in 2018 daarom AI-technieken ingezet om begrippen uit verschillende domeinen te analyseren én daarmee de resultaten te verfijnen.

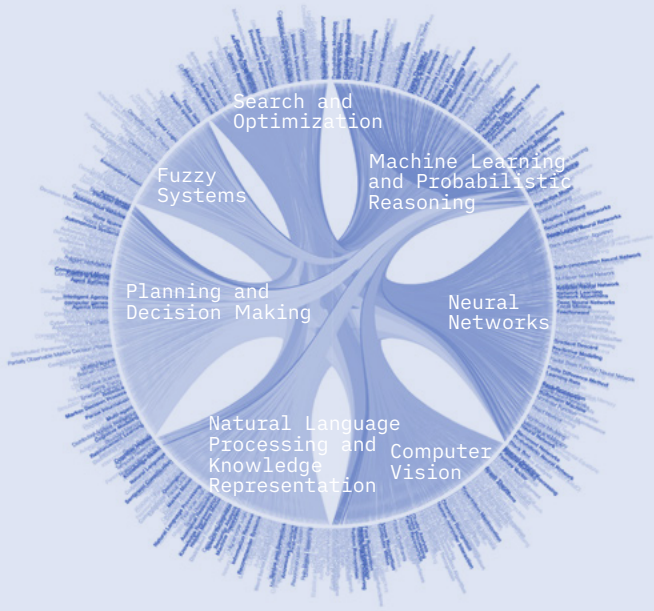
Allereerst zijn relevante boeken, wetenschappelijke literatuur, online cursussen (MOOCS), patentaanvragen en nieuwsberichten geanalyseerd door gebruik te maken van onder andere Natural Language Processing ↴ (NLP). Op basis van deze inzichten is een shortlist samengesteld van 800 *keywords*. Vervolgens zijn met behulp van deze keywords 6 miljoen publicaties geselecteerd uit Scopus, de database van Elsevier. Deze resultaten zijn echter niet representatief aangezien er ook documenten in de selectie zitten die weliswaar de begrippen bevatten, maar deze niet in een AI-context hanteren (zoals neural networks binnen de biologie). Er is daarom een supervised machine learning ↴-techniek ingezet om de keywords te classificeren, variërend van een hoge, gemiddelde of lage relevantie. Op basis van deze classificatie zijn uiteindelijk 600.000 publicaties geselecteerd die het wetenschappelijke corpus van AI vormen.

Op basis van de relevante keywords en publicaties zijn vervolgens met behulp van een unsupervised machine

learning ↴-techniek clusters gemaakt. Hieruit zijn zeven dominante AI-onderzoeksgebieden voortgekomen:

> Search and Optimization
> Fuzzy Systems
> Natural Language Processing and Knowledge Representation
> Computer Vision
> Machine Learning and Probabilistic Reasoning
> Planning and Decision Making
> Neural Networks

Onderzoek in Machine Learning, Neural Networks en Computer Vision ↴ kennen het grootste volume aan onderzoeksoutput en -groei. De clustering geeft tevens een kijkje in de verschillende domeinen. Zo zijn de keywords vanuit de industrie enorm verspreid over de verschillende onderzoeksgebieden, terwijl er hele specifieke keywords vanuit de wetenschap verschijnen in Neural Networks, keywords vanuit het onderwijs in Search and Optimization en keywords vanuit de media in Planning and Decision Making. Globaal gezien komen de minste keywords voort uit het mediadomein.



‘The AI research field clusters around seven main research areas’. – Bron: Elsevier Analytical Services (2018)

Opvallend is dat er relatief weinig begrippen refereren aan ethische vraagstukken. Ondanks de aanwezigheid van AI in ons dagelijks leven en het groeiende besef dat AI mogelijk ethische dilemma’s gaat veroorzaken, is ethiek in relatie tot AI volgens de studie van Elsevier momenteel nog een relatief klein onderzoeksgebied.

AI en de wisdom of the crowd

Een clustering van relevante onderzoeksgebieden geeft weliswaar een beter inzicht in de complexiteit van AI, maar een breder begrip van de woorden ontbreekt. Het bevat geen sentiment; het laat niet zien wat mensen van belangrijke keywords vinden en of ze positief of negatief tegenover de begrippen staan. In dit geval kan *wisdom of the crowd* uitkomst bieden. IBM ontwikkelde de Project Debater, een AI-systeem dat in 10 minuten 300 miljoen artikelen kan analyseren én op basis hiervan argumenten kan formuleren voor of tegen stellingen. Het systeem neemt het inmiddels op tegen professionele debaters. Het doel volgens IBM is om mensen te helpen bij het opbouwen van krachtige argumenten en het nemen van goed geïnformeerde besluiten. Dit systeem is tevens ingezet om het Project Debater – Speech by Crowd platform te ontwikkelen. Hiermee kunnen *crowdsourced* besluiten worden genomen. Op basis van opinies kan het systeem een overtuigende uiteenzetting geven van argumenten en zowel een pleidooi voor als tegen de stelling geven. Allereerst verzamelt het systeem de argumenten. Deze worden vervolgens geclassificeerd op basis van sentiment (voor of tegen) en beoordeeld op kracht (sterk of zwak). Overtollige begrippen worden vervolgens gefilterd en de argumenten worden daarna geclusterd. Op basis van deze clusters wordt aan de hand van een standaard structuur een narratief opgesteld. Een van de crowdsourced stellingen gaat over de vraag of de ontwikkeling van zelfrijdende auto’s moet worden stopgezet.

We should stop the development of autonomous cars

35% pro

The topic of autonomous cars are complex and sometimes leads to tough discussions, for example when considering whether We should stop the development of autonomous cars. The following is an attempt to lay out a few major arguments in favor.

> Next, I will discuss three issues which explain why We should stop the development of autonomous cars. I will begin by claiming that autonomous car can cause accidents. I will also discuss issues related to the claim that self-driving cars introduce new, poorly understood safety risks. And finally I will elaborate on how autonomous cars will cause many types of jobs to disappear.

About accidents. Autonomous cars will not have the capacity to determine what is the lesser evil in terms of an accident that can cause property damage or result in injuries to humans or animals. Until there is a legal framework covering insurance of and therefore remuneration of individuals affected by an accident involving autonomous cars, the development of autonomous cars should be stopped. There are no opportunities for a autonomous car to be distracted, which is a leading cause of collisions in the United States at present. Of the 62 accidents in California involving self-driving cars in autonomous mode over the last four years, the autonomous vehicle was only responsible in one case. Autonomous cars can be hacked and can cause massive crashes that would take thousands of lives.

> I also mentioned safety. Self-driving cars can be hacked just as any other computing device, which can compromise the safety of the riders in a variety of ways. Autonomous cars will create safety and liability issues as they cannot be 100% precise and accurate. Software updates to autonomous cars will have undesirable side effects on safety. Two recent deaths involving Uber and Tesla vehicles using driverless cars have raised the debate on safety to levels that threaten to significantly delay or derail adoption of the technology. Autonomous cars are a security risk because if they can be hacked they can be potentially weapons of destruction. They are a massive security risk since the technology necessary can be hacked and used to orchestrate large scale attacks.

The last issue I mentioned was jobs. Autonomous cars will destroy the driver's ed industry and DMVs, putting hundreds of people out of jobs. They will cause many types of jobs to disappear -- taxi drivers, truckers, car dealers, driving schools and many more. They will cause many people to loose their jobs.

> That concludes my speech. I thank you for your time and wish you a safe drive home.

65% Con

The topic of autonomous cars are controversial and sometimes leads to complex conversations, like when examining whether We should stop the development of autonomous cars. This text will consider a few major arguments against.

> Next, I will discuss three issues which explain why We should continue the development of autonomous cars. I will try to convey that autonomous cars will greatly expand the freedom of elderly people. I will also show that autonomous cars will actually help alleviate traffic and road. And lastly I will talk about how autonomous cars are an important technology to improve safety on the road.

Starting with old age. Autonomous cars would be so useful for the many, and increasing number, of elderly and disabled persons who cannot drive. Autonomous cars

allows old people, kids, teenager and handicapped people to be mobile and thus extend their quality of life. Autonomous cars will help people who can't drive, like children or elderly people. Autonomous driving will support the continued participation in the community of older people who live in remote communities and find driving too demanding as they age. Autonomous vehicles have the potential to enhance the mobility and independence of the disabled and elderly populations.

Turning to roads. Autonomous cars will actually help alleviate traffic and road clog by forcing all cars to maintain a consistent speed and prevent man-made problems with merging and road rage that cause traffic. Autonomous cars will make roads safer because the cars will drive slower to precisely follow all traffic rules. Autonomous cars will increase national security by allowing the state to better monitor the roads. Modern cars are parked 96% of the time. Autonomous vehicles and car sharing would require less space on the road and free up more space for parks, recreational centres and housing in populated cities. Because people will need to look less for parking spaces, and since they can drop you off at your destination, find a parking spot after, and then pick you up when you're done, then clearly autonomous cars eliminate problems with finding parking.

Finally, technology. Development of autonomous cars will also boost the development of electric cars, since those two technologies are closely related. Autonomous cars are an important technology to improve safety on the road.

I thank you for your time. Drive safely.

Voordelen en beperkingen

AI is in staat om in een korte tijd een grote hoeveelheid data te analyseren. Het classificeren en clusteren van dermate grote hoeveelheden publicaties is handmatig onbegonnen werk. Dominante begrippen in onderzoek geven aan waar de aandacht van wetenschappers naar uitgaat en waar hoogstwaarschijnlijk de doorbraken en toepassingen van de toekomst op gebaseerd zullen zijn. Door een breder begrip van taal aan de hand van Natural Language Processing kan informatie tevens voorzien worden van sentiment en geeft het een beter inzicht in gedachten en opvattingen.

» AI in research is like Ctrl F on steroids. «

De inzet van AI bij onderzoek kan dus een krachtige tool zijn om informatie te classificeren en te clusteren. Toch sluipen er een aantal foutjes in. Zo staat er in het pleidooi van IBM's Project Debater vóór het stopzetten van ontwikkelingen van de zelfrijdende auto een tegenargument (bij 'about accidents'). Het interpreteren van informatie blijft vooralsnog mensenwerk. Zeker wanneer het gaat om uitspraken over de toekomst. AI blijft een geavanceerde vorm van statistiek ↴. Je kunt er veel mee automatiseren, maar de basisassumptie is dat de toekomst net zo zal zijn als het verleden. De data wordt immers uit dezelfde kansverdeling getrokken.

AI in research is like Ctrl F on steroids ↵

Cartoon over de data-afhankelijkheid van machine learning – Bron: SMBC by Zach Weinersmith

AI is in staat om automatisch patronen te herkennen. Als je het specifiek genoeg maakt kun je daarmee voorspellen. Zo kun je een bepaald patroon uit het verleden vertalen naar het heden en extrapoleren naar de toekomst. Huizenprijzen laten zich bijvoorbeeld prima voorspellen als je de juiste parameters hebt. De toekomst van AI gaat echter over grote fundamentele veranderingen, waarbij dergelijke methoden niet goed werken. AI heeft nog geen theoretisch inzicht in hoe de wereld werkt en kan daardoor geen uiteenlopende toekomstscenario's schetsen voor dergelijke complexe onderwerpen.

» Men heeft de komst van deep learning ook niet zien aankomen door naar de papers van de voorgaande 10 jaar te kijken. «

-- Maarten Stol, BrainCreators

Met een combinatie van menselijk inzicht, analytics en machine learning kun je wel betere voorspellingen doen. Ook hier zie je dat het beroep van bijvoorbeeld een toekomstonderzoeker niet zomaar verdwijnt, maar dat het takenpakket wel degelijk verandert.

Conclusies en vooruitblik

De toekomst van AI hangt af van verschillende factoren en laat zich niet eenvoudig voorspellen. Het is niet alleen afhankelijk van technologische doorbraken, maar ook van heersende politieke en economische belangen én de kennis die we hebben over ons brein. Want voorlopig zijn we nog niet in staat om de wonderen van menselijke intelligentie volledig te ontrafelen. De verwachting is dat AI ons daar wel steeds meer bij zal gaan helpen. Het kan patronen in data vinden die wij zelf niet zouden zien. AI heeft eerder al planeten ontdekt die wij over het hoofd gezien hebben. Het is dan niet ondenkbaar dat het op vergelijkbare wijze met nieuwe inzichten komt over ons brein. Een handige tool is het sowieso. Want handmatig alle informatie doornemen is met de toenemende hoeveelheid data steeds minder haalbaar. IBM's Poject Debater kan in 10 minuten 300 miljoen artikelen analyseren. Een vergelijkbare taak zou voor een mens meer dan 2000 jaar kosten.

Wanneer we menselijke intelligentie precies gaan benaderen blijft echter de vraag. Volgens sommige experts zullen we general AI al in 2029 bereiken, volgens anderen duurt het nog tot 2200. Dit neemt echter niet weg dat de meesten er wel in geloven dát het bereikt zal worden. Zoals Alan Turing in 1950 al voorspelde zal de doorbraak liggen in de benadering van hoe kinderen leren, namelijk een combinatie van kijken en praten. Hiervoor zullen de twee kampen van deep learning en formal logic elkaar moeten gaan vinden en samenwerken. Samenwerking lijkt sowieso het codewoord voor de toekomst van AI. Verschillende vakgebieden moeten gezamenlijk AI-systemen gaan bestuderen. AI-systemen zullen namelijk steeds meer autonoom acties ondernemen en keuzes maken. Dit maakt dat ze steeds meer als volwaardige spelers in de samenleving zullen gaan functioneren. Het gaat dus niet alleen over de technologische werking, maar ook over de psychologische, sociologische, economische en ethische consequenties.

Deel II van deze reeks gaat daarom over de impact van AI. Welke sociaal-culturele implicaties heeft de ontwikkelingsrichting van AI voor besluitvorming en welke mogelijke toekomstscenario's kunnen hierdoor uiteengezet worden? Er zijn namelijk verschillende toekomstbeelden mogelijk. Ook hier spelen verschillende belangen een rol. Wie gaat de machtsstrijd om werelddominantie in AI winnen, China of Amerika? Of toch Google of Amazon? Het hangt daarbij sterk samen met de relatie met de technologie die we voor ogen hebben. Moet AI menselijke arbeid gaan vervangen of gaat het ons ondersteunen? Worden het autonome wezens waarmee we samenleven of blijven het tools die wij bedienen? Dit hangt nauw samen met de stip die we op de horizon zetten. Blijft dit menselijke intelligentie, of benaderen we artificiële intelligentie als een nieuwe op zichzelf staande vorm van intelligentie?

Aan de hand van de mogelijke toekomstscenario's kunnen we onszelf vervolgens de vraag stellen welk scenario (of welke combinatie van elementen uit de verschillende scenario's) het meest wenselijk is. In deel III gaan we daarom in op ethische vraagstukken die een rol spelen bij de impact van AI op besluitvorming in de toekomst.

De computer als medemens

Door Frank van Harmelen, Professor Knowledge Representation & Reasoning, Vrije Universiteit Amsterdam

Elke krantenlezer weet het: we zitten midden in de AI-revolutie. De beschreven AI-technieken hebben één ding gemeen: ze werken in wat ik noem ‘vervangingsmodus’. Ze zijn allemaal bedoeld om uiteindelijk de mens te vervangen. Kan een computer beter vertalen? Dan hebben we die tolk niet meer nodig.

Dit perspectief van AI als ‘De Grote Vervanger’ is onaantrekkelijk en onverstandig om allerlei redenen. Ten eerste zijn er de sociaal-maatschappelijke redenen. Verregaande automatisering en verregaande verstoring van de arbeidsmarkt is geen aanlokkelijk perspectief. Zelfs als we iedereen van een basisinkomen voorzien, dan nog zouden we aan veel mensen de betekenis ontnemen die werk hen geeft. En net als met eerdere grote revoluties op de arbeidsmarkt zullen de lusten en de lasten van zo'n ‘vooruitgang’ ongelijk verdeeld zijn: meestal worden de rijken rijker. Daarnaast zijn er belangrijke technische en wetenschappelijke redenen. In het AI-onderzoek komen we er steeds meer achter dat AI heel anders is dan menselijke intelligentie. In sommige zaken zijn computers veel sterker dan mensen: computers hebben een perfect geheugen, ze kunnen patronen zien die voor het menselijk oog onzichtbaar zijn en ze kunnen redeneringen volgen die veel langer zijn dan voor het menselijk brein mogelijk is. Maar aan de andere kant zijn mensen zich subtiel bewust van de context waarin ze zich bevinden, ze beseffen welke handelingen in zo'n context wel of niet sociaal of zelfs moreel-ethisch acceptabel zijn en anders dan computers zijn mensen bij uitstek samenwerkers.

Gegeven deze complementariteit tussen artificiële en menselijke intelligentie ligt het helemaal niet zo voor de hand om de een door de ander te vervangen. Het is dan veel logischer om teams te maken waarin de twee vormen van intelligentie samenwerken, zodat ze elkaar aanvullen en samen beter presteren dan elk van hen apart zou doen. In plaats van AI noemen we dat ‘hybride intelligentie’. Om zulke teams van samenwerkende mensen en intelligente computers te bouwen moet er nog veel gebeuren. Alle huidige en toekomstige doorbraken van het AI-onderzoek zullen nodig zijn, maar daarboven moeten we ook aandacht gaan besteden aan vaardigheden voor computers waar tot nu toe nauwelijks over nagedacht werd. Die vaardigheden laten zich samenvatten in het acroniem ‘CARE’: Collaborative, Adaptive, Responsible en Explainable.

Collaborative: Intelligente computers moeten leren samenwerken. Ze moeten zich bewust zijn van wat de menselijke ‘collega's’ wel en niet weten, wat hun intenties en doelen zijn en moeten daarop in kunnen spelen. *Adaptive:* De wereld waarin wij leven en werken is open, veranderlijk en vaak onvoorspelbaar. Het zal dus niet langer volstaan om een computer te trainen voor een bepaalde taak. In plaats daarvan moeten computers permanent de wereld observeren, zien wat er verandert en leren om zich daaraan aan te passen. *Responsible:* Bij echte hybride intelligentie zullen computers niet alleen feitelijke, maar ook morele afwegingen moeten kunnen meenemen in hun handelen. *Explainable:* Een computer moet kunnen uitleggen waarom ze een bepaalde beslissing neemt. De huidige deep-learning systemen zijn nog een ‘black box’: deze zal open moeten.

Dit zijn allemaal grote wetenschappelijke uitdagingen voor het AI-onderzoek, maar ze zullen leiden tot hybride teams waarin mensen en computers samenwerken en daarmee beter presteren dan elk van hen apart. Op weg naar de computer als medemens?

Waarom zouden we menselijke intelligentie nastreven met AI?

door Rudy van Belkom

In de zoektocht naar intelligente machines staat het benaderen, of zelfs overstijgen, van menselijke intelligentie al vanaf de jaren ’50 als prominente stip op de horizon. Naast de verschillende technologische uitdagingen, is deze zoektocht volgens mij om drie redenen enorm lastig.

1: We weten niet precies hoe intelligentie bij mensen werkt.
2: We hebben voor veel relevante begrippen (zoals bewustzijn) geen algemeen geaccepteerde operationele definitie, waardoor het bestaan ervan moeilijk aan te tonen is.
3: We verleggen continu wat we intelligent vinden.

Dat alles maakt het lastig om intelligentie ‘te klonen’. Experts zijn het er dan ook niet over eens wanneer we human-level AI gaan bereiken. De stip op de horizon verschuift met de tijd mee en lijkt continu even ver weg te staan. Toch is het niet ondenkbaar dat ‘de intelligentiecode’ wordt gekraakt. Voor schaken leek ooit een vorm van menselijke intelligentie nodig te zijn; je moest strategisch kunnen denken en je tegenstander kunnen inschatten. Inmiddels weten we dat alle ‘what ifs’ geprogrammeerd kunnen worden en een abstracte representatie en brute rekenkracht voldoende zijn om de mens te verslaan. Wat als taken die nu heel complex lijken ook met een relatief eenvoudig algoritme op te lossen zijn? In dat opzicht lijkt creativiteit het nieuwe schaken. AI is nu al in staat om kunstwerken te maken en muziekstukken te componeren. Veel mensen hebben moeite om dit te accepteren als échte creativiteit. Hier speelt tevens de filosofische discussie of wij zelf ook niet gewoon geprogrammeerd zijn en dus niet zo autonoom handelen als we denken. AI kan daarmee de volgende ‘belediging’ voor de mensheid worden. Machines waren ons eerder al de baas in fysieke arbeid en rekenkracht en nu staat ons intellectueel vermogen ook al op het spel. Dit vermogen heeft ons altijd onderscheiden van alle andere wezens op aarde en heeft ons (in ieder geval gevoelsmatig) controle gegeven over onze omgeving. Het is dan ook niet vreemd dat sommige mensen zich tegen AI afzetten. AI is hoe dan ook een spiegel voor de mensheid. Het leert ons enorm veel over onszelf en stelt ons fundamentele vragen over het menszijn.

Maar de hamvraag is wat mij betreft of we überhaupt menselijke intelligentie moeten nastreven. Duikboten zwemmen niet zoals vissen en vliegtuigen vliegen niet zoals vogels; waarom zouden computers dan wel moeten denken als mensen? Wanneer je een spin menselijke intelligentie geeft dan zal deze zich niet als een mens gaan gedragen, maar als een ‘superspin’ die nog betere webben kan spinnen en prooien kan vangen. We gaan pas echt stappen zetten als we het idee loslaten dat we superieure wezens zijn. De mens is niet superieur aan insecten, we zijn allebei geëvolueerd voor andere doeleinden. Mensen hebben weliswaar meer geavanceerde cognitieve vaardigheden, maar insecten kunnen hoogst-waarschijnlijk een nucleaire ramp overleven. ‘Succes’ is dus contextafhankelijk en daardoor relatief.

We moeten onszelf gaan afvragen voor welk doel we intelligente machines willen inzetten, in plaats van het als doel op zichzelf te zien. Hoe kunnen we met behulp van intelligente machines een betere wereld creëren? En wat is een betere wereld precies? Mens en machine zouden wat mij betreft moeten samenwerken vanuit hun eigen specialisaties. Laat complexe statistiek over aan computers en sociaal gevoelige vraagstukken aan mensen. Waarom zouden we emoties inbouwen in machines? Ik zou er juist naar streven om computers zo objectief mogelijk te maken. Daar zijn mensen met alle evolutionair geprogrammeerde vooroordelen en emoties namelijk helemaal niet zo goed in.

Begrippenlijst

Affective Computing

Affective Computing gaat over systemen die emoties kunnen opsporen en herkennen. 1.1 Hoe werkt AI? ↴

Afhankelijk zijn van data

Een van de grootste beperkingen van AI is dat het afhankelijk is van grote hoeveelheden data. Bij deep learning wordt daarom ook wel gesproken over data-hungry neural networks. Dit maakt dat de technologie niet goed is in randgevallen, waarbij weinig data beschikbaar is. 2.3 Voordelen en beperkingen ↴

AI

De meest dominante associatie met AI is machine learning. Uit onderzoek van de World Intellectual Property Organisation (WIPO) uit 2019 blijkt tevens dat machine learning de meest dominante AI-technologie is die in patentaanvragen is opgenomen. Binnen deze toekomstverkenning richten we ons daarom hoofdzakelijk op machine learning. 1.1 Hoe werkt AI? ↴

AI-winters

AI-winters zijn perioden waarin de interesse in en financiering voor AI-onderzoek laag is. 1.2 Do(n't) believe the hype ↴

Algoritmen

Een algoritme is een wiskundige formule. Het is een eindige reeks instructies die vanuit een gegeven begintoestand naar een vooraf bepaald doel leidt. 1.1 Hoe werkt AI? ↴

Anders denken

Een mens kan goed kijken, maar niet zo goed redeneren. Een computer kan goed redeneren, maar niet zo goed kijken. Juist wat wij intuïtief doen en weinig energie kost, is voor een computer complex en kost veel energie. En vice versa. Dit wordt ook wel Moravec's paradox genoemd. 2.1 Perspectieven op intelligentie ↴

Beeldherkenningssytemen

Bij toepassingen op het gebied van beeldherkenning wordt vaak gebruik gemaakt van een zogenaamd Convolutional Neural Network (CNN). CNN werkt als een filter die over een afbeelding beweegt en op zoek gaat naar de aanwezigheid van bepaalde eigenschappen. 1.1 Hoe werkt AI? ↴

Begrip

Zelfs wanneer je alle kennis over de wereld in een computer programmeert, dan blijft de vraag of een computer daadwerkelijk begrip heeft van zijn handelen. Dit begrip wordt ook wel intentionaliteit genoemd. 2.1 Perspectieven op intelligentie ↴

Besluitvorming

We willen graag geloven dat wij mensen rationele wezens zijn. Maar het besluitvormingsproces is grillig. Beeldvorming en ambities spelen naast feitelijke kennis een belangrijke rol. 2.2 Geautomatiseerde besluitvorming ↴

Bewustzijn

Bewustzijn maakt dat we ons onder andere bewust zijn van onze emoties en gedachten, dat we keuzes en de gevolgen daarvan kunnen ervaren en dat we kunnen reflecteren op ons gedrag. 2.1 Perspectieven op intelligentie ↴

Biases

Aangezien we evolutionair zijn geprogrammeerd om zoveel mogelijk energie te besparen sluipen er wel wat 'denkfouten' in onze informatieverwerking. Deze denkfouten worden ook wel cognitive biases genoemd. 2.2 Geautomatiseerde besluitvorming ↴

Clusters

Het algoritme gaat dan zelfstandig op zoek naar overeenkomsten in de data en probeert op deze wijze patronen te herkennen. 1.1 Hoe werkt AI? ↴

Common Sense

Common sense bestaat uit alle kennis over de wereld; van fysieke en zichtbare aspecten, tot culturele en dus meer impliciete regels, zoals hoe je met elkaar omgaat. 2.1 Perspectieven op intelligentie ↴

Computer Vision

Computer Vision gaat over systemen die in staat zijn om informatie uit beelden te halen en deze zo te herkennen. 1.1 Hoe werkt AI? ↴

Convolutional Neural Network

Een Convolutional Neural Network werkt als een filter die over een afbeelding beweegt en op zoek gaat naar de aanwezigheid van bepaalde eigenschappen.

Deep learning

Deep learning is een machine learning-methode die gebruik maakt van verschillende gelaagde artificial neural networks. 1.1 Hoe werkt AI? ↴

Deep reasoning

Combinatie van deep learning en symbolic AI. Neurale netwerken zorgen ervoor dat het systeem op basis van patroonherkenning kan ‘zien’ en de symbolische programmering zorgt ervoor dat het systeem kan ‘redeneren’. 3.1 Expert-voorspellingen ↴

Deepfakes

Met behulp van AI kun je een persoon dingen laten zeggen of doen die hij of zij in werkelijkheid nooit gezegd of gedaan heeft. Deze techniek wordt ook wel deepfake genoemd, een combinatie van de woorden deep learning en fake. Een bekend voorbeeld is de gemanipuleerde video van Barack Obama. 2.3 Voordelen en beperkingen ↴

Definitie van AI

“*AI zijn intelligente systemen die zelfstandig taken kunnen uitvoeren in complexe omgevingen en eigen pre-staties kunnen verbeteren door te leren van ervaringen.*”, aldus De Nationale AI-cursus. 1.1 Hoe werkt AI? ↴

Eerste fase van AI

In de eerste fase van AI, van 1957 tot eind jaren ‘90, werd voornamelijk gebruikt gemaakt van formal logic, oftewel als-dan regels. Deze vorm van AI was vooral gericht op high level cognition, zoals redenering en probleemoplossing. 1.1 Hoe werkt AI? ↴

Empathie

Empathie is de vaardigheid om je in te kunnen leven in de situatie en gevoelens van anderen. Door empathie zijn we in staat om ook de non-verbale communicatie van anderen te lezen en begrijpen. 2.1 Perspectieven op intelligentie ↴

Experts

Futurist Martin Ford interviewde voor zijn boek *Architects of Intelligence* 18 wereldwijd gerespecteerde experts op het gebied van AI. De voorspellingen wanneer general AI bereikt zal worden lopen uiteen van 2029 (11 jaar vanaf 2018) tot 2200 (182 jaar vanaf 2018). 3.1 Expert-voorspellingen ↴

Expertsystemen

Een expertsysteem, of kennisgebaseerd systeem, gebruikt geprogram-meerde kennis van menselijke experts om een specifiek probleem op te lossen. 1.1 Hoe werkt AI? ↴

Explainibility

Het wordt steeds lastiger voor mensen om te achterhalen op basis van welke data de uitkomsten van AI gebaseerd zijn. AI wordt daarom nu nog vaak vergeleken met een black box. 2.2 Geautomatiseerde besluit-vorming ↴

Formele logica

In de eerste fase van AI, van 1957 tot eind jaren ‘90, werd voornamelijk gebruikt gemaakt van formal logic, oftewel als-dan regels. Deze vorm van AI was vooral gericht op high level cognition, zoals redenering en probleemoplossing. 1.1 Hoe werkt AI? ↴

General AI

Artificial General Intelligence zou in staat moeten zijn om alle intellectuele taken uit te voeren die een mens ook kan uitvoeren. 1.1 Hoe werkt AI? ↴

Human level AI

Oftewel Artificial General Intelligence (AGI). Deze vorm van intelligentie zou in staat moeten zijn om alle intellectuele taken uit te voeren die een mens ook kan uitvoeren. 1.1 Hoe werkt AI? ↴

Intelligentie

Intelligentie kan omschreven worden als een aaneenschakeling van verstandelijke vermogens, processen en vaardigheden, zoals kunnen redeneren en je flexibel kunnen aanpassen in nieuwe situaties. 2.1 Perspectieven op intelligentie ↴

Intentionaliteit

Zelfs wanneer je alle kennis over de wereld in een computer programmeert, dan blijft de vraag of een computer daadwerkelijk begrip heeft van zijn handelen. Dit begrip wordt ook wel intentionaliteit genoemd. 2.1 Perspectieven op intelligentie ↴

Kennis over ons brein

Neurowetenschappers zijn steeds beter in staat om bepaalde functies aan specifieke hersengebieden te koppelen. Toch weten we over de werking van intelligentie nog relatief weinig. 2.1 Perspectieven op intelligentie ↴

Kennissystemen

Een kennisgebaseerd systeem, of expertsysteem, gebruikt geprogrammeerde kennis van menselijke experts om een specifiek probleem op te lossen. 1.1 Hoe werkt AI? ↴

Leerervaringen

Er wordt onderscheid gemaakt tussen drie vormen van leren: supervised learning, reinforcement learning en unsupervised learning. 1.1 Hoe werkt AI? ↴

Machine Learning

Het gaat bij machine learning over een revolutie waarin mensen niet meer programmeren (als dit, dan dat), maar waarin machines zelf regels afleiden uit data. 1.1 Hoe werkt AI? ↴

Motorkap

We kunnen de complexe concepten die wij aan het brein willen koppelen, zoals bewustzijn en vrije wil, nog niet modelleren en we kunnen ze niet een-op-een linken aan gebieden in het brein. 2.1 Perspectieven op intelligentie ↴

Narrow AI

Artificial Narrow Intelligence is een vorm van AI die zeer goed is in het doen van specifieke taken. Denk hierbij aan schaken, aanbevelingen doen en het geven van kwantificeerbare voorspellingen. 1.1 Hoe werkt AI? ↴

Natural Language Processing

Natural Language Processing gaat over de vaardigheid van een systeem om menselijke taal te begrijpen. Toepassingen vind je in chatbots, vertaalmachines, spamfilters, automatisch gegeneerde samenvattingen en zoekmachines. 1.1 Hoe werkt AI? ↴

Neurale netwerken

Artificiële neurale netwerken worden gebruikt binnen deep learning en zijn oorspronkelijk gebaseerd op het menselijk brein, waarbij neuronen laagsgewijs met elkaar verbonden zijn. 1.1 Hoe werkt AI? ↴

Recurrent Neural Networks

Recurrent Neural Networks zijn in staat om hun interne status, oftewel geheugen, te gebruiken om reeksen van invoer te verwerken (zoals spraak en video). 1.1 Hoe werkt AI? ↴

Reinforcement Learning

Reinforcement Learning gaat over leren door middel van trial and error. De voorbeelden worden hierbij niet vooraf gelabeld, maar het algoritme krijgt op basis van de resultaten achteraf feedback. 1.1 Hoe werkt AI? ↴

Statistiek

AI is in de basis ‘gewoon’ wiskunde. Weliswaar een enorm geavanceerde vorm van wiskunde, maar het blijft wiskunde. Het is boven alles een middel om een optimalisatiedoel te bereiken. 1.1 Hoe werkt AI? ↴

Strenger voor technologie

Uit onderzoek van onder andere de University of Penssylvania uit 2014 blijkt dat wanneer mensen een algoritme een kleine en betekenisloze fout zien maken, dat de kans dan groot is dat het vertrouwen volledig verloren gaat. Dit wordt door onderzoekers ook wel algorithm aversion genoemd. 2.2 Geautomatiseerde besluitvorming ↴

Superintelligence

Artificial Super Intelligence (ASI) kan bereikt worden wanneer AI het kunnen van het menselijk brein op alle mogelijke domeinen overstijgt. 1.1 Hoe werkt AI? ↴

Supervised (machine) learning

Supervised Learning gaat over het zorgvuldig samenstellen en vooraf labelen van trainingsdata. Het algoritme leert dan data te classificeren aan de hand van gelabelde voorbeelden. 1.1 Hoe werkt AI? ↴

Symbol grounding

Symbol grounding is het gegeven dat Intelligentie ontstaat door interactie met de omgeving. Representaties in onze geest krijgen pas betekenis door de koppeling met sensomotorische ervaringen. 2.1 Perspectieven op intelligentie ↴

Systeem 1

Systeem 1 staat voor het snelle, automatische denken. Het gebeurt onbewust en wordt gekoppeld aan instinctieve en emotionele keuzes. Het is intuïtief en kost ons weinig energie. 2.1 Perspectieven op intelligentie ↴

Systeem 2

Systeem 2 staat voor het trage, beredeneerde denken. Het is een bewust proces en wordt gekoppeld aan rationele keuzes. Informatie wordt verzameld en afgewogen en de verwerking ervan kost veel energie. 2.1 Perspectieven op intelligentie ↴

Trainen

Er wordt onderscheid gemaakt tussen drie vormen van leren: supervised learning, reinforcement learning en unsupervised learning. 1.1 Hoe werkt AI? ↴

Transformaties

Bij transformatie worden de beperkingen van denkstijlen doorbroken. Dit wordt ook wel gezien als de diepste vorm van creativiteit. 2.1 Perspectieven op intelligentie ↴

Unsupervised (machine) learning

Unsupervised Learning gaat over algoritmen die leren om zelf naar clusters te zoeken in grote hoeveelheden ongelabelde data. Het algoritme gaat dan zelfstandig op zoek naar overeenkomsten in de data en probeert op deze wijze patronen te herkennen. 1.1 Hoe werkt AI? ↴

Vertrouwen in technologie

Volgens de Gartner Hype Cycle zijn er bij introductie van een nieuwe technologie hoge verwachtingen, tot de eerste failures zich aandienen. Hierna kantelen de verwachtingen en het vertrouwen verdampt. De ontwikkelingen gaan echter door, de marktintroductie wordt doorgezet en de verwachtingen stabiliseren waarna we de technologie accepteren. 1.2 Do(n’t) believe the hype ↴

Bronnen

AI IMPACTS (2015). AI Timeline Surveys. Geraadpleegd van https://aiimpacts.org/ai-timeline-surveys/

AINED (2018). AI voor Nederland. Geraadpleegd van https://www.vno-ncw.nl/sites/default/files/aivnl_20181106_0.pdf

Araujo, T. et al. (2018). Automated Decision-Making Fairness in an AI-driven World: Public Perceptions, Hopes and Concerns. Geraadpleegd van http://www.digicomlab.eu/wp-content/uploads/2018/09/20180925_ADMbyAI.pdf

Armstrong, S. (2014). The errors, insights and lessons of famous AI predictions – and what they mean for the future. Geraadpleegd van http://www.fhi.ox.ac.uk/wp-content/uploads/FAIC.pdf

Benitez-Quiroz, C.F, Srinivasan, R. & Martinez, A.M. (2019). Facial color is an efficient mechanism to visually transmit emotion. Geraadpleegd van https://www.pnas.org/content/115/14/3581

Bloom, P. (2018). Against Empathy; the case for rational compassion. New York: Vintage Books.

Bloom, N. et al. (2019). Are Ideas Getting Harder to Find? Geraadpleegd van https://web.stanford.edu/~chadj/IdeaPF.pdf

BNC (2018). Mededeling Kunstmatige Intelligentie voor Europa. Geraadpleegd van https://www.rijksoverheid.nl/documenten/kamerstukken/2018/06/01/fiche-7-mededeling-kunstmatige-intelligentie-voor-europa

Boden, M.A. (2004). Creativity in a Nutshell. Geraadpleegd van https://www.researchgate.net/publication/209436199_Creativity_in_a_nutshell

Bostrom, N. (2016). Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press.

Brooks, R.A. (1990). Elephants Don't Play Chess. Geraadpleegd van http://people.csail.mit.edu/brooks/papers/elephants.pdf

Cialdini, R.M. (2007). Influence: The Psychology of Persuasion. Geraadpleegd van http://elibrary.bsu.az/books_400/N_232.pdf

Clarke, A.C. (2000). Profiles of the future; An Inquiry into the Limits of the Possible. London: Phoenix.

Carroll, L. & Mondrick, M.E. (1976). Racial Bias in the Decision to Grant Parole. Geraadpleegd van https://www.jstor.org/stable/3053205?seq=1#page_scan_tab_contents

Datillo, A. et al. (2019). Identifying Exoplanets with Deep Learning II: Two New Super-Earths Uncovered by a Neural Network in K2 Data. Geraadpleegd van https://www.cfa.harvard.edu/~avanderb/Deep_Learning_2.pdf

De Nationale AI-cursus (2019, 8 april). Wat is AI? Geraadpleegd van https://app.ai-cursus.nl/programmas/de-nationale-ai-cursus/wat-is-ai/start

Denkwerk (2019) Arbeid in transitie. Geraadpleegd van https://denkwerk.online/media/1048/denkwerk_arbeidintransitie_hoge_resolutie.pdf

Dietvorst, B., Simmons, J.P. & Massey, C. (2014). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. Geraadpleegd van https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2466040

Digital Communication Methods Lab (2018). Automated Decision-Making Fairness in an AI-driven World: Public Perceptions, Hopes and Concerns. Geraadpleegd van http://www.digicomlab.eu/wp-content/uploads/2018/09/20180925_ADMbyAI.pdf

Edmonds, E. (2019). Three in Four Americans Remain Afraid of Fully Self-Driving Vehicles. Geraadpleegd van https://newsroom.aaa.com/2019/03/americans-fear-self-driving-cars-survey/

Elsevier (2018). ArtificiaI Intelligence: How knowledge is created, transferred, and used. Geraadpleegd van https://www.elsevier.com/__data/assets/pdf_file/0010/823654/ACAD-RL-AS-RE-ai-report-WEB.pdf

Callahan, M. (2019). It’s time to correct neuroscience myths. Geraadpleegd van https://cos.northeastern.edu/news/its-time-to-correct-neuroscience-myths/

Frey, C.B. & Osborne, M.A. (2013). The Future of Employment. Geraadpleegd van https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf

Ford, M. (2018) Architects of Intelligence. Birmingham: Packt Publishing.

Gold, J. & Shadlen, M. (2002). Banburismus and the Brain: Decoding the Relationship between Sensory Stimuli, Decisions, and Reward. Geraadpleegd van https://www.sciencedirect.com/science/article/pii/S0896627302009716

Hao, K. (2019). We analyzed 16,625 papers to figure out where AI is headed next. https://www.technologyreview.com/s/612768/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/

Hawkins, J. (2003). How brain science will change computing. Geraadpleegd van https://www.ted.com/talks/jeff_hawkins_on_how_brain_science_will_change_computing/transcript

High-Level Expert Group on Artificial Intelligence (2019). Ethics guidelines for trustworthy AI. Geraadpleegd van https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

IRENA (2018). Renewable Energy and Jobs – Annual Review 2018. Geraadpleegd van https://irena.org/publications/2018/May/Renewable-Energy-and-Jobs-Annual-Review-2018

Kahneman, D. (1979). Prospect Theory: An Analysis of Decision under Risk. Geraadpleegd van https://www.uzh.ch/cmsssl/suz/dam/jcr:00000000-64a0-5b1c-0000-00003b7ec704/10.05-kahneman-tversky-79.pdf

Kahneman, D. (2012). Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.

Keijzer, M.C.G. (2018). Geraadpleegd van https://www.wrr.nl/onderwerpen/artificiele-intelligentie/documenten/brieven/2018/06/11/adviesaanvraag-artificial-intelligence

Kshirsagar, A. (2019). Monetary-Incentive Competition Between Humans and Robots: Experimental Results. Geraadpleegd van https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3323698

KPMG (2018). Ready, Set, Fail?: Avoiding setbacks in the intelligent automation race. Geraadpleegd van https://advisory.kpmg.us/content/advisory/en/index/articles/2018/new-study-findings-read-ready-set-fail.html

Martin, E. (2018). Moore’s Law is Alive and Well. Geraadpleegd van https://medium.com/predict/moores-law-is-alive-and-well-adc010ea7a63

MCC Ventures (2019). The state of AI: Divergence. Geraadpleegd van https://www.mmcventures.com/wp-content/uploads/2019/02/The-State-of-AI-2019-Divergence.pdf

McCarthy, J. (1955). A proposal for the Dartmouth summer research project on Artificial Intelligence. Geraadpleegd van http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

Microsoft (2019). Artificial Intelligence Report: Outlook for 2019 and Beyond. Geraadpleegd van https://pulse.microsoft.com/uploads/prod/2018/10/WE_AI_Report_2018.pdf

Muller, C. (2017). Kunstmatige intelligentie - De gevolgen van kunstmatige intelligentie voor de (digitale) eengemaakte markt, de productie, consumptie, werkgelegenheid en samenleving. Geraadpleegd van https://www.robertwent.nl/wp-content/uploads/sites/16/2018/05/EESC-over-KI-nl.pdf

OECD (2018). Automation, skills use and training. Geraadpleegd van https://www.oecd-ilibrary.org/docserver/2e2f4eea-en.pdf?expires=1556107382&id=id&accname=guest&checksum=8B1776A7B-0324F91865DF602DAE2489E

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. Geraadpleegd van https://stevenpinker.com/publications/cognitive-niche-coevolution-intelligence-sociality-and-language

PwC (2019). Exploiting the AI revolution. Geraadpleegd van https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

ProPublica (2016). Machine Bias. Geraadpleegd van https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Rahwan, I. (2019). Machine behavior. Geraadpleegd van https://www.nature.com/articles/s41586-019-1138-y

Russel, S.J. & Norvig, P. (2010). Artificial Intelligence: a modern approach. Geraadpleegd van https://faculty.psau.edu.sa/filedownload/doc-7-pdf-a154ffbcec538a4161a406abf62f5b76-original.pdf

Singh, S (2018). Cousins of Artificial Intelligence. Geraadpleegd van https://towardsdatascience.com/cousins-of-artificial-intelligence-dda4edc27b55

StartupDelta (2018). Artificial Intelligence The Netherlands: Startup Report. Geraadpleegd van http://www.startupdelta.org/reports/

Tenenbaum, J. (2019). The neuro-symbolic concept learner: interpreting scenes, words and sentences from natural supervision. Geraadpleegd van https://openreview.net/pdf?id=rJgMlhRctm

Turing, A.M. (1950). Computing Machinery and Intelligence. Geraadpleegd van http://cogprints.org/499/1/turing.html

Verberne, F. (2015). Trusting a virtual driver Similarity as a trust cue. Geraadpleegd van http://alexandria.tue.nl/extra2/794808.pdf

Verheij, B. (2018). Arguments for good artificial intelligence. Geraadpleegd van http://www.ai.rug.nl/~verheij/publications/oratie/oratie_Bart_Verheij.pdf

Vis, B. (2017). Waarom politicologen betwijfelen of burgers zouden moeten stemmen. Geraadpleegd van https://www.brainwash.nl/bijdrage/waarom-politicologen-betwijfelen-of-burgers-zouden-moeten-stemmen

Wang, Y. & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Geraadpleegd van https://psyarxiv.com/hv28a/

Wilson, B. et al (2019). Predictive Inequity in Object Detection. Geraadpleegd van https://arxiv.org/abs/1902.11097

WIPO (2019). WIPO Technology Trends 2019: Artificial Intelligence. Geraadpleegd van https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf

Wissner-Gross, A. (2013). A new equation for intelligence. Geraadpleegd van https://www.ted.com/talks/alex_wissner_gross_a_new_equation_for_intelligence

Zong, S, et al. (2019). Analyzing the Perceived Severity of Cybersecurity Threats Reported on Social Media. Geraadpleegd van https://arxiv.org/pdf/1902.10680.pdf

Bijlagen

Over de auteur

Rudy van Belkom MA studeerde Brand, Design & Reputation Management aan het European Institute for Brandmanagement (EURIB). Voor zijn masterthesis deed hij onderzoek naar de factoren die doorslaggevend zijn bij de acceptatie van radicale technologische innovaties. Vertrouwen in de technologie blijkt hierin een cruciale rol te spelen. Tot zijn indiensttreding bij STT deed hij onder andere onderzoek naar de rol van sciencefiction bij toekomstonderzoek in het Fontys lectoraat Futures Research & Trendwatching en gaf hij jarenlang les in conceptontwikkeling en trendonderzoek in het hoger onderwijs.

Deelnemers

Voor deze toekomstverkenning zijn verschillende experts geraadpleegd. Namens STT en mijzelf enorm bedankt voor het delen van alle kennis en inspiratie:

Paul Blank*	NWO	Program Officer TTW
Sanne Blauw	De Correspondent	Correspondent Ontcijferen
Marc Burger	Capgemini	CEO
Taco Cohen*	Qualcomm AI Research	Deep Learning Researcher
Sybo Dijkstra	Philips	Head of Data Strategy & AI
Patrick van der Duin*	STT	Directeur
Erik Fledderus	SURF	Algemeen Directeur
Bernard ter Haar	Ministerie SZW	DG Sociale zekerheid en Integratie
Frank van Harmelen	VU	Professor Knowledge Representation & Reasoning
Emy Hermens	CQM	Data Science Consultant
Fred Herrebout	T-Mobile	Senior Strategy Manager
Michiel van den Hauten*	Belasting-dienst	Directeur Innovatie en strategie
Jeroen Heres	Ministerie J&V	Hoofd centrale eenheid strategie
Noor Huijboom	Ministerie BZK	Policy Coordinator
Peter Hulsen*	CQM	Partner
Marijn Janssen	TU Delft	Professor ICT & Governance
Leon Kester*	TNO	Senior Research Scientist
Maria de Kleijn-Lloyd*	Elsevier	SVP Analytical Services
Meine Koeslag	ISPT	Program Officer Industry
Maarten Kruithof*	TNO	Researcher Deep learning
Daan Kolkman*	JADS	Onderzoeker besluitvorming en AI
Mignon Lebon	Ministerie EZ&K	Directie Digitale Economie
Stefan Leijnen*	The Asimov Institute	Director
Carien Leushuis	CQM	Data Science & Optimization Consultant
Marike van Lier Lels	Commissaris	RELX, NS Groep en Dura Vermeer
Max van der Linden*	UvA	Docent psychologie
Steven Luijtjens	Ministerie BZK	Voormalig directeur DIO
Leendert van Maanen*	UvA	Assistant professor Psychological Methods Department
Marieke van Putten*	Ministerie BZK	Senior Innovation

Max Remerie*	Independent	Adviser & interim manager
Maarten de Rijke	UvA	Professor of AI and Information Retrieval
Jelmer de Ronde*	SURF	Projectmanger SURFnet
Klamer Schutte*	TNO	Lead Scientist Intelligent Imaging
Haroon Seikh	WRR	Wetenschappelijk medewerker
Gerard Smit	IBM	CTO
Nanda van der Stap	TNO	Expertise group for intelligent imaging
Maarten Stol*	Brain-Creators	Principal Scientific Adviser
Benjamin Timmermans	IBM	Center for Advanced Studies Benelux
Barbara Vis*	Utrecht University	Professor Politics & Governance
Jasper Wognum	Brain-Creators	CEO

* deze experts hebben de teksten van deze publicatie voorzien van feedback

Stichting Toekomstbeeld der Techniek

De Stichting Toekomstbeeld der Techniek (STT) is in 1968 opgericht door het Koninklijk Instituut Van Ingenieurs (KIVI). STT is een onafhankelijke stichting die gefinancierd wordt uit bijdragen van overheid en bedrijfsleven. STT voert brede toekomstverkenningen uit op het snijvlak van technologie en samenleving die domein-overstijgend en interdisciplinair zijn. Het Algemeen Bestuur (AB) van STT bestaat uit topmensen uit de overheid, het bedrijfsleven, de onderzoekswereld en uit maatschappelijke organisaties. Het AB denkt mee over de STT-programmering, is betrokken bij verkenningen en vormt een belangrijke denktank waarbinnen de bestuursleden praten over toekomstige technologische ontwikkelingen en innovatie.

Daarnaast ontplooit de STT Academy diverse activiteiten, zoals het co-financieren van bijzondere leerstoelen, methodiekontwikkeling en het beheer van het Netwerk Toekomstverkenningen en Young STT. Deze laatste bestaat uit young high potentials uit de deelnemende organisaties.

Informatie over STT, haar activiteiten en haar producten is te vinden op www.stt.nl.

Eerder verschenen publicaties STT

STT90 *Bioprinters, grondstof-rotondes en brainternet? Hoe wij produceren, consumeren en herverdelen in 2050.* Silke den Hartog-de Wilde, 2019

STT89 *Veiligheid in de toekomst.* Carlijn Naber, 2019

STT88 *Het eeuwige leren. Over leren, technologie en de toekomst.* Dhoya Snijders, 2018

STT87 *En toen ging het licht aan…; Transities naar een emissievrij energiesysteem.* Soledad van Eijk, 2017

STT86 *Data is macht. Over Big Data en de toekomst.* Dhoya Snijders, 2017

Colofon

STT92 deel I: Duikboten zwemmen niet

Onderzoek en projectleiding: Rudy van Belkom
Tekst- en taalredactie: Japke Schreuders
Grafisch ontwerp: YURR studio, Yurr Rozenberg

NUR-nr. 950

Trefwoorden
Technologie, artificiële intelligentie, toekomst, Nederland, forecasting

In deze reeks over de toekomst van AI verschijnen tevens deel II: de impact van AI en deel III: de ethiek van AI.

U kunt dit werk toeschrijven aan Stichting Toekomstbeeld der Techniek/ Rudy van Belkom, 2019.
Stichting Toekomstbeeld der Techniek
Koninginnegracht 19
2514 AB Den Haag
070-302 98 30
info@stt.nl www.stt.nl

» Machines have calculations. We have understanding. Machines have instructions. We have purpose. There's one thing only a human can do. That's dream. So let us dream big. «

-- Garry Kasparov, 2017