

Slim, slimmer, slimst



New Scientist
Pocket Science

Hoe kunstmatige
intelligentie de
mens een
turboboost geeft

Bennie Mols

Inhoud

Inleiding 7

1. Goddelijke zet 13
2. Het geheim van de kast 25
3. Kleine beetjes menselijkheid 36
4. Psychopiloot 49
5. Hybride intelligentie 55
6. Eurekaboost 66
7. Creativiteitsboost 78
8. Superintelligentie 92
9. Een Babelvis voor iedereen 100

Verantwoording en dankwoord 109

Index 111

Meer lezen? 117

Inleiding

Tussen 2003 en 2008 volgde ik elke zomer drie weken lang een taal cursus Pools. Mijn vrouw is Poolse, we komen vaak in Polen, dus is het handig om de taal enigszins te beheersen. Aan die cursus heb ik Poolse vrienden en vriendinnen overgehouden met wie ik af en toe in hun taal chat of mail. Praten gaat me op huis-tuin-en-keukenniveau redelijk af, maar schrijven is andere koek. In de jaren dat ik de taal cursussen volgde, bestonden er nog geen vertaalmachines die me konden helpen. Ik was veel tijd kwijt aan het zo correct mogelijk proberen te schrijven, altijd met een woordenboek en grammaticalijsten bij de hand. Het schrijven ging langzaam, zo frustrerend langzaam dat ik het steeds vaker helemaal achterwege liet.

Toen werd in 2006 Google Translate gelanceerd. In de beginjaren werd daar nogal lacherig over gedaan, want de machine maakte vaak hilarische fouten. Fokke en Sukke zeiden in 2011 in de cartoon *Fokke en Sukke doen alles met Google Translate*: ‘Dames en heren! Zijn we verheugd te kunnen aankondigen dat het nu taalhelp online lastige taalproblemen die blijven tot het verleden horen verwacht.’

Enkele jaren na mijn laatste cursus was Google Translate voor mij al zo goed dat ik mijn vrienden weer met plezier in

het Pools kon schrijven. Sindsdien combineer ik mijn beperkte kennis van het Pools en de juist veel grotere kennis van de wereld, met de grotere woordenschat, de betere grammatica, de betere spelling en de veel hogere snelheid van de vertaalmachine. Zo verbaas ik mijn Poolse vrienden keer op keer. De vertaling van Google Translate en mij samen is veel beter dan die van mijzelf, en ook beter dan die van de vertaalmachine in zijn eentje. Af en toe zie ik dat de vertaalmachine niet begrijpt wat hij vertaalt, of dat hij bij dubbelzinnige woorden de mist in gaat en dan corrigeer ik de automatische vertaling.

In ruim tien jaar tijd heeft Google Translate, en meer recent ook de in Duitsland ontwikkelde evenknie DeepL, zich ontwikkeld tot een krachtige, razendsnelle supervertaler. Niet foutloos, en zonder begrip van wat woorden in de tastbare wereld betekenen, maar toch superpraktisch. In een fractie van een seconde vertaalt de machine stukken tekst in meer dan honderd talen. Ondanks de fouten kunnen we ineens communiceren met mensen wier taal we niet of nauwelijks spreken. Ineens kunnen we teksten lezen in talen waarvan we niets begrijpen. Ineens kunnen we onze eigen teksten veel beter en sneller vertalen. We zijn het volstrekt normaal gaan vinden dat machines dit kunnen. Of, zoals een van de pioniers van het vakgebied kunstmatige intelligentie (in het Engels *artificial intelligence*, AI) John McCarthy het zei: ‘Zodra het werkt, noemt niemand het meer AI.’

In de afgelopen jaren is AI een begrip geworden. Kunstmatige intelligentie is nu gewoon AI en AI is steeds meer in ons alledaagse leven doorgedrongen. Zonder dat we het nog in de gaten hebben is AI standaard in onze computers, tablets, smartphones en digitale diensten gaan huizen. Kunstmatige intelligentie vertaalt onze teksten, filtert spam uit onze

e-mails, ondertitelt automatisch onze video's, zet ons gesproken woord om in tekst, herkent gezichten in onze foto's, geeft pratend antwoord op onze vragen en beveelt nieuwe producten, films en boeken aan op basis van onze eerdere aankopen.

Steeds vaker ondersteunt AI artsen bij het stellen van diagnoses, helpt het bij juridische beslissingen, overheidsbeslissingen en fraudedetectie, en bepaalt het of we op een vliegveld wel dezelfde persoon zijn als die op de pasfoto. AI zorgt voor een explosie van nieuwe robottoepassingen en maakt auto's steeds meer autonoom. De mogelijkheden zijn eindeloos. We zijn nog maar net begonnen ze te ontdekken.

De Brits-Amerikaanse AI-onderzoeker en ondernemer Andrew Ng ziet kunstmatige intelligentie als de nieuwe elektriciteit. 'Net zoals elektriciteit honderd jaar geleden bijna alles transformeerde, zal AI de komende jaren vrijwel elke denkbare bedrijfstak transformeren.' AI is inderdaad bijna net zo onzichtbaar en onmisbaar aan het worden als elektriciteit.

We hebben het nu al enkele bladzijden over AI, maar wat is dat precies?

De wetenschap van de AI bestudeert machines die dingen doen die intelligentie vereisen wanneer mensen ze zouden doen. Dat is een handige definitie die de moeilijke filosofische vraag wat intelligentie is, bewust uit de weg gaat. De *Encyclopaedia Britannica* hanteert als definitie van intelligentie: 'Menselijke intelligentie bestaat uit de vermogens om te leren van ervaringen, zich aan te passen aan nieuwe situaties, abstracte concepten te begrijpen en te hanteren en kennis te gebruiken om de omgeving te veranderen.'

Wie wat beter kijkt naar het begrip intelligentie ziet al snel dat er verschillende soorten bestaan: cognitieve (redeneren), sociale (verbale en non-verbale communicatie), emotionele

2

Het geheim van de kast

‘Zelfs de meest geavanceerde kunstmatige intelligentie is nog lang niet in staat tot het oplossen van problemen die vierjarige kinderen met gemak oplossen.’

– Alison Gopnik (psycholoog)

Een man met een stapel papier in zijn handen loopt naar een gesloten kast toe. In de hoek van de kamer kijkt een klein jongetje toe. De man botst tegen de kast op, doet een paar stappen terug, loopt weer naar voren en botst voor de tweede keer tegen de kast op. Weer doet hij een paar stappen terug. Plotseling loopt het jongetje naar de kast, opent de deuren en kijkt omhoog naar de man. De man loopt weer naar voren en terwijl het jongetje omlaag kijkt naar een boekenplank, legt de man de stapel papier in de kast.

Het jongetje is anderhalf jaar oud en heeft deze situatie nooit eerder gezien. Toch begrijpt hij de bedoeling van de man en de functie van de kast. Ook is hij in staat naar de kast toe te lopen, de deurhendel beet te pakken en de deur te openen. Met zijn lichaamstaal communiceert hij zelfs met de man.

Ondanks meer dan een halve eeuw aan AI-ontwikkeling, ondanks een decennialange exponentiële groei in computer-

kracht, kan geen enkele robot wat dit jongetje kan en heeft geen enkel AI-systeem het begripsniveau van dit jongetje.

Gezond verstand

De scène van het jongetje en de man met de stapel papier komt uit een reeks psychologische experimenten waarin onderzoekers altruïsme bij kinderen bestudeerden. Tegelijk is het een perfecte illustratie van hoeveel begrip een kind van anderhalf jaar al heeft. En het feit dat geen enkel AI-systeem dit kan, toont hoeveel AI nog heeft te leren van menselijke intelligentie.

Wat het kind onderscheidt van zelfs de beste AI-systemen is wat wel 'gezond verstand' heet, impliciete kennis of alledaagse kennis. Dit is kennis die we van jongs af aan spelenderwijs opdoen, die niemand ons expliciet heeft verteld en die we nooit op internet hebben opgezocht. Wanneer ik een glas water in mijn hand pak en dat op een harde ondergrond laat vallen, weet ieder kind dat het glas in stukken breekt en het water over de vloer wegloopt. In tegenstelling tot AI-systemen leren kinderen bovendien volledig uit zichzelf. Ze hebben een intrinsieke motivatie om de wereld om hen heen spelenderwijs te exploreren en doen dat ook nog eens in een sociale en culturele context.

Begrijpen wat de bedoeling is van de man met de stapel papier, is een voorbeeld van intuïtieve psychologie. Intuïtieve psychologie gaat over het begrijpen en voorspellen van gevoelens en emoties, het begrijpen van de bedoelingen van andere mensen en het je kunnen verplaatsen in de gedachten en gevoelens van andere mensen. Drie maanden oude baby's weten al dat mensen hun doelen meestal met zo min mogelijk inspanning proberen te bereiken. Tien maanden oude baby's kunnen al een intuïtieve inschatting maken hoeveel

mensen ervoor over hebben om een bepaald doel te bereiken, bijvoorbeeld hoe graag ze iets willen opeten.

Onderdeel van deze intuïtieve psychologie is ook wat in de filosofie *theory of mind* heet. Dit gaat over je kunnen indenken wat een ander denkt, gelooft, weet en bedoelt. In sociale interactie gebruiken mensen dat voortdurend. Ik weet dat jij weet dat Amsterdam de hoofdstad is van Nederland. Maar ik weet ook dat jij weet dat ik weet dat Amsterdam de hoofdstad is van Nederland.

Mensen ontwikkelen die vaardigheid geleidelijk. Ergens tussen het derde en vierde levensjaar leren kinderen begrijpen wat iemand anders wel of niet weet. En zo rond de tienjarige leeftijd begrijpen kinderen wat het betekent als iemand anders iets expliciet belooft. Nog later leren kinderen redeneren over wat ze denken dat anderen over hen denken.

Begrijpen dat een glas dat uit iemands handen glipt, naar beneden valt, is een voorbeeld van intuïtieve natuurkunde. In de eerste twaalf maanden leren baby's bijvoorbeeld dat voorwerpen niet zomaar ophouden te bestaan, niet zomaar naar een andere plek teleporteren en dat ze niet door obstakels heen kunnen. Het kiekeboe-spel (waarin iemand zich achter zijn handen, een deur of een boom verbergt, en na een paar seconden weer tevoorschijn komt en dan 'kiekeboe' roept) is een mooi voorbeeld van deze ontluikende kinderwereld. Baby's leren ook al vroeg een begrip van de driedimensionale ruimte ontwikkelen. Ze leren de concepten zwaartekracht, traagheid en oorzaak en gevolg. Allemaal voorbeelden van intuïtieve natuurkunde: het begrijpen hoe alledaagse voorwerpen zich gedragen onder invloed van de natuurkrachten.

Intuïtieve biologie, ten slotte, gaat over het begrijpen wat levende wezens onderscheidt van niet-levende dingen. Een kind dat een hond ziet, weet dat de hond leeft, dat hij beweegt, dat hij eten en drinken nodig heeft en dat hij een

eigen wil heeft. Een beeldherkennend AI-systeem dat heeft geleerd om een hond te herkennen op een foto heeft geen idee wat die hond wel en niet allemaal kan en wil.

Bij elkaar genomen vormen intuïtieve psychologie, intuïtieve natuurkunde en intuïtieve biologie dat wat we gezond verstand noemen. Een van de grootste wetenschappelijke uitdagingen voor AI is om computers ook gezond verstand te geven. Cognitiewetenschapper Gary Marcus noemt gezond verstand ‘*dé* voorwaarde voor algemene intelligentie’. Marcus: ‘Tot het moment dat machines gezond verstand hebben, zitten we vast aan een gespecialiseerde kunstmatige intelligentie, die zelden robuust is en nooit zo flexibel als het menselijk verstand.’

Leren van weinig voorbeelden

De *deep learning*-techniek die we in het vorige hoofdstuk kort bespraken, blinkt uit in patroonherkenning, maar heeft daar wel veel meer voorbeelden voor nodig dan de patroonherkenning van het menselijk brein. Bovendien is menselijke intelligentie veel meer dan patroonherkenning. Menselijke intelligentie gaat ook over het modelleren van de wereld. Bij dat modelleren van de wereld hoort verklaren en begrijpen wat we zien, verbeelden wat we zouden kunnen zien, maar nog niet hebben gezien, problemen oplossen, acties plannen en nieuwe modellen ontwikkelen terwijl we meer over de wereld leren.

Gezond verstand stelt ons in staat modellen van de wereld te ontwikkelen en zo ook van slechts een beperkt aantal voorbeelden te leren. Een baby kan het gezicht van zijn moeder of vader herkennen nadat hij hen slechts een paar keer heeft gezien. AI die leert van slechts enkele voorbeelden is een grote wetenschappelijke uitdaging.

Een Babelvis voor iedereen

‘Zelfs als we de machines onder controle kunnen houden, bijvoorbeeld door de stroom op strategische momenten uit te schakelen, moeten we ons als soort zeer nederig voelen.’

– Alan Turing (computerpionier), 1951

‘De Babelvis is klein, geel en bloedzuigerachtig, en waarschijnlijk het vreemdste ding in het heelal... als je een Babelvis in je oor steekt kun je direct alles begrijpen wat er tegen je gezegd wordt in welke taal dan ook.’

Zo fantaseert de schrijver Douglas Adams over een universele vertaalmachine in het even hilarische als geniale science-fictionboek *The Hitchhiker’s Guide to the Galaxy*. Toen het boek in 1979 verscheen was de Babelvis pure fantasie. ‘Maar nu is de Babelvis binnen handbereik’, zegt de 85-jarige Indiaas-Amerikaanse AI-pionier Raj Reddy wanneer ik hem in de herfst van 2022 spreek. ‘Allerlei taaltechnologieën, zoals automatische spraakherkenning en automatische vertaling, zijn nu goed genoeg om te worden gebruikt.’

Reddy werkt in de AI-wetenschap sinds begin jaren zestig, enkele jaren na de geboorte van AI als vakgebied. Ik vraag hem hoe hij aankijkt tegen de AI-ontwikkelingen van de afgelopen tien jaar, waarin machinaal leren een spectaculaire

ontwikkeling heeft doorgemaakt. Reddy: ‘Ik werk al bijna zestig jaar op dit terrein en ik had niet gedacht dat ik dergelijke taaltechnologieën nog zou meemaken. Een belangrijke toepassing is het helpen van de mensen onderaan de maatschappelijke ladder. Er zijn ongeveer twee miljard mensen in de wereld die niet kunnen lezen of schrijven. Dankzij AI-systemen kan over tien jaar zelfs een analfabeet elk boek lezen, elke film bekijken en met iedereen, waar ook ter wereld, een gesprek voeren in zijn of haar moedertaal.’

Ik vertel deze anekdote omdat er in de afgelopen jaren vooral in de media en de politiek zoveel nadruk is komen te liggen op de risico’s van AI-technologieën dat we soms lijken te vergeten hoeveel nieuwe kansen ze bieden. Een Babelvis voor iedereen is echte vooruitgang, net zoals antibiotica, vaccins, computers en het internet echte vooruitgang hebben opgeleverd.

Ogenschijnlijk triviale technologieën kunnen grote positieve sociale gevolgen hebben. Zo vertelt de Zweedse arts en hoogleraar internationale gezondheid Hans Rosling in zijn boek *Feitenkennis* over de impact die het kopen van een wasmachine begin jaren vijftig van de vorige eeuw had bij hem thuis: ‘Nou Hans, we hebben de machine geladen,’ zei zijn moeder. ‘De machine gaat het werk doen. Dus nu kunnen wij naar de bibliotheek.’ Rosling vervolgt: ‘De was ging erin, boeken kwamen eruit. Dank je wel, industrialisatie, dank, staalfabriek, dank, energiecentrale, dank chemische industrie, omdat jullie ons de tijd hebben gegeven om boeken te lezen.’

Rosling geeft in zijn boek tien hoofdredenen waarom het beter in de wereld gaat dan we vaak denken. Op basis van wetenschappelijke data laat hij zien dat het deel van de wereldbevolking dat in extreme armoede leeft in de afgelopen twintig jaar is gehalveerd, hoe sterk de gemiddelde

levensverwachting is gestegen (tot zeventig jaar), hoe de meeste kinderen op de wereld worden gevaccineerd (tachtig procent), hoe sterk het aantal mensen dat toegang heeft tot elektriciteit is gestegen (tachtig procent) en hoe technologie op talloze terreinen heeft bijgedragen aan een betere wereld.

Soms lijkt het dat hoe welvarender, gezonder en langer we leven, hoe minder risico's we accepteren. Maar innoveren zonder risico's te nemen gaat niet; een zekere mate van risico's zullen we moeten accepteren.

AI-risico's

Wat zijn dan de specifieke risico's van AI? De lijst is lang: AI kan fouten maken, AI kan vooringenomen zijn, AI kan banen bedreigen, AI-beslissingen kunnen onuitlegbaar zijn, AI kan gebruikt worden om desinformatie te creëren, mensen kunnen vaardigheden kwijtraken, mensen kunnen te veel op AI vertrouwen, de samenwerking tussen mens en AI kan verkeerd gaan, en AI kan voor machtsverschuivingen in de maatschappij zorgen. En dit is nog maar een selectie.

Omdat AI al zoveel toepassingen heeft gekregen en er nog veel meer gaat krijgen, is het noodzakelijk om die risico's te onderzoeken en zoveel mogelijk proberen te beperken. Dat gebeurt dan ook volop. Zo hebben tal van organisaties in de afgelopen jaren ethische principes opgesteld, waaronder de Verenigde Naties en het Europese Parlement. Het aantal opgestelde principes varieert sterk, maar wat al deze initiatieven met elkaar gemeen hebben valt samen te vatten in drie overkoepelende ethische principes: verantwoordingsplicht, transparantie en verantwoordelijkheid.

Verantwoordingsplicht heeft betrekking op de beslissingen die het AI-systeem neemt. Die moeten uit te leggen en te rechtvaardigen zijn tegenover de mensen waarmee het

systeem communiceert. Om de verantwoordingsplicht te waarborgen, moeten besluiten kunnen worden afgeleid en verklaard uit de gebruikte algoritmen.

Transparantie heeft betrekking op het systeem als geheel. Volgens welke motieven, waarden en principes is het systeem ontworpen? Wat zijn de mogelijkheden en de beperkingen van het systeem?

Verantwoordelijkheid staat voor het belangrijkste principe: mensen zijn verantwoordelijk voor wat de machines doen, en niet de machines. Die mensen kunnen ontwikkelaars zijn, bouwers, eigenaars of gebruikers. Verantwoordelijkheid houdt ook in dat mensen ervoor moeten zorgen dat kunstmatige intelligentie iedereen ten goede komt en niet alleen de *happy few*.

Wet- en regelgeving

Ethische principes alleen zijn echter onvoldoende om AI-risico's te beperken. Uiteindelijk moeten wet- en regelgeving het juiste gedrag ook afdwingen. Zo stelde de Europese Commissie in 2021 een voorstel op voor een *AI Act* die tot doel heeft een gemeenschappelijk regelgevend en juridisch kader voor kunstmatige intelligentie in te voeren. Een definitief akkoord wordt in de loop van 2023 verwacht.

De voorgestelde *AI Act* deelt AI-toepassingen in drie risicocategorieën in. In de eerste categorie vallen toepassingen met een onaanvaardbaar risico, zoals een sociaal kredietsysteem waarin het gedrag van burgers op basis van een uitgebreide dataverzameling en AI-analyse wordt beoordeeld (China heeft een vorm van een sociaal kredietsysteem geïntroduceerd). Deze toepassingen worden binnen de EU verboden. In de tweede categorie vallen toepassingen met een hoog risico, zoals een AI-systeem dat cv's van sollici-