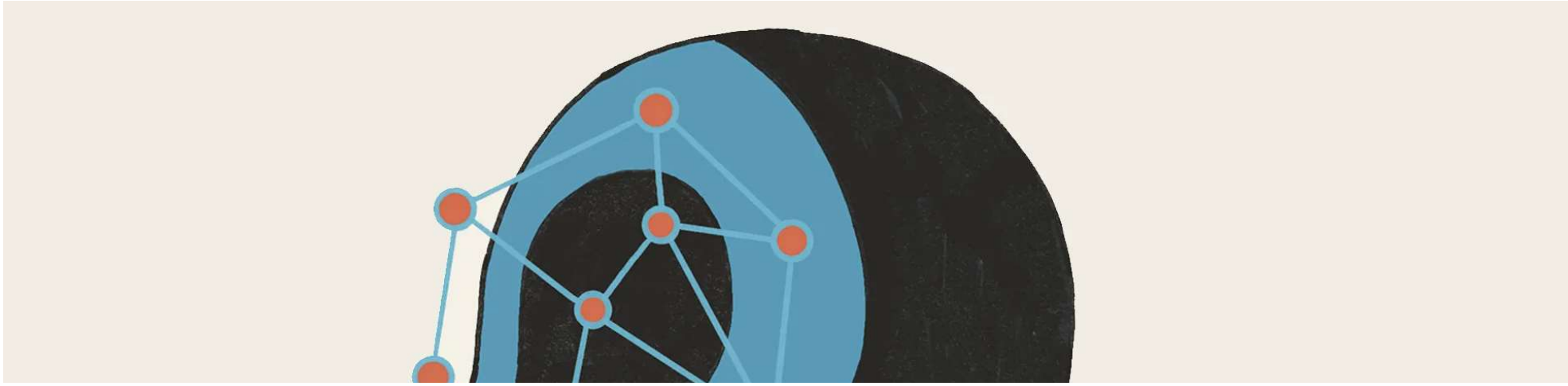
N.B. Het kan zijn dat elementen ontbreken aan deze printversie.

Luister naar
16:40

# AI zit verblufte, betoverde én ontnuchterde filosofen op de hielen

**Filosofie** Denken over kunstmatige intelligentie is meer dan alleen maar kluiven op abstracte vragen. Is het walhalla nabij, of het armageddon? Steeds meer filosofen houden zich met AI bezig.

Sjoerd de Jong 22 februari 2023 Leestijd 9 minuten



Onlangs doopte de Duitse *Süddeutsche Zeitung* kunstmatige intelligentie de vierde narcistische „krenking" voor de mensheid. Eerst was er Copernicus (uit het centrum van het heelal gegooid), toen kwam Darwin (we stammen af van de apen), daarna Freud (ten prooi aan onbewuste driften) - en nu worden we ook nog ingehaald door machines die ons het laatste privilege afnemen: intelligentie. Adieu oude mens, uiteengespat in een confetti van bits en bytes, vrolijk nagezwaaid door ChatGPT.

Dat is een schrikbeeld voor cultuurpessimisten en complotdenkers die een 'transhumane' toekomst met angst en walging tegemoet zien. Het is de opmaat naar een door „financiële opperheren" gecontroleerde „slavenmaatschappij", jammert het tweewekelijkse paranoia-blad *Gezond Verstand*, dat zich ook afzet tegen alles wat met 'transgenderisme' te maken heeft. Tegenover hen staan de hallelujaroepers die AI zien als een stralende toekomst voor de mensheid, zoals in *Homo Deus* van pop-filosoof Yuval Harari.

Of is kunstmatige intelligentie - saaier, maar wel zo interessant - geen van beide, niet het virtuele walhalla maar ook geen digitaal armageddon? Hoe denken filosofen daarover?

Vakfilosofen die zich bezighouden met AI - en hun aantal groeit snel - zijn allang druk met het 'demystificeren' van AI, het temperen van zowel overdreven optimisme als het zwartste pessimisme. Ja, de technologie rukt in razend tempo op, maar

verschilt nog steeds van menselijke intelligentie. Inderdaad, er dreigen gevaren, maar vooral aan de 'menselijke' kant. Nadenken over 'verantwoordelijke' AI als antwoord daarop is *hot* in de Academie. Het moet 'moralisering' van AI opleveren, toegesneden op de menselijke maat, niet achteraf maar al in de ontwerpfase. Ingenieurs en ethici trekken dan samen op in het laboratorium.

## DYNAMIEK VAN DE DIALOOG

Maar eerst de schok van het nieuwe. Jan Broersen, sinds vier jaar hoogleraar logische methoden en AI in Utrecht, wil aan het begin van het gesprek wel kwijt dat hij „zwaar onder de indruk" is van de ChatGPT - inclusief de filosofische én sociale implicaties ervan. „Het heeft de natuurlijkheid die je gewend bent van mensen en dieren." Ook frappant: de nieuwe chatbot, getraind op het platform Reddit, wéét wanneer het moet stoppen met antwoorden, een cruciale eigenschap. Broersen: „Het kent de dynamiek van de dialoog."

Maar hoe goed is het? Inmiddels is het een wereldwijd gezelschapsspel geworden om de praatgrage bot op blunders of onzin te betrappen - er zijn al talloze voorbeelden te vinden, van onjuistheden tot verzinsels en verkeerde voorspellingen. Maar die feilbaarheid doet nog niet af aan de invloed ervan, zegt Broersen. „Je hoort op verjaardagen ineens vragen over bewustzijn, intelligentie en autonomie die altijd vooral aan de universiteit werden behandeld. Dat is een effect van dit type AI: het heeft klassieke filosofische kwesties gedemocratiseerd."

Lees een interview met Jan Broersen:
**'Er blijft in AI toch altijd iets ontbreken'**

De eerste vraag die Broersen aan de chatbot stelde was dan ook „of hij dacht de Turing-test te kunnen doorstaan”. De beroemde wiskundige Alan Turing (1912-1954) geldt als grondlegger van de moderne computerwetenschap. Zijn werk riep de vraag op of je van een rekenmachine die geprogrammeerde instructies volgt kunt zeggen dat hij een spel als schaken echt ‘begrijpt’.

En? Broersen, lachend: „Het antwoord was nee.” Van een student had hij het fout gerekend, zegt hij. ChatGPT lijkt het spel van vragen en antwoorden immers prima te begrijpen. Het punt is, zegt Broersen, hij begrijpt zichzélf niet. „Ook de beste AI-taalprogramma’s zijn niet meer dan probabilistische rekenmonsters. Ze kunnen woordvolgorde verbijsterend goed voorspellen, maar betekenis en logische structuur ontgaan ze.” ChatGPT is al een „stochastische papegaai” genoemd: een ijverige naprater. Broersen: „Je denkt dat je een stem hoort, maar in feite hoor je een kakofonie van miljoenen stemmen, een mix van de megadata. Een spreker, maar geen persoon.’

Dat hoeft niet te verbazen. Taalfilosofen spreken van syntaxis, semantiek en pragmatiek. Woorden krijgen betekenis in een context: die van een zin, opmerking of betoog, intenties van sprekers en ‘taalhandelingen’ (beweren, beloven, voorspellen, zweren, en andere). Machines weten daar slecht raad mee, zeker met dubbele bodems, metaforen of ironie. Iets vergelijkbaars geldt voor moraal, zegt Broersen. Je kunt AI-machines zoals zelfrijdende

auto's leren moreel te handelen, maar dat blijft „een morele schil" aan de buitenkant. Een ethische *governor* die 'bovenop' het model wordt aangebracht. „Het zit niet diep genoeg."

> *Ook de beste AI-taalprogramma's zijn niet meer dan probabilistische rekenmonsters*
>
> **Jan Broersen** –hoogleraar logische methoden en AI

En zoals Broersen intussen over de grootste 'narcistische krenking' heen. Maar de Utrechtse filosoof maakt zich wél grote zorgen over de maatschappelijke effecten van ChatGPT. Hij heeft een hele lijst: nog sterkere machtsconcentratie bij de grote techbedrijven, proliferatie van schijnkennis, wegdrukken of verdwijnen van kleine talen, uitvergroten van cognitieve verschillen tussen mensen en van economische ongelijkheid. „De chatbot is nu nog gratis, maar iedereen die er geld in heeft gestopt zal een *return on investment* willen zien." Lachend: „Eerst maken ze ons verslaafd en dan laten ze ons betalen." Broersen is dan ook blij dat er regelgeving komt zoals die van de EU, die 'verantwoorde AI' nastreeft.

Die regulering komt dan nog van buiten, in de wetenschappelijke wereld zelf buigen zich ook steeds meer filosofen over de vraag hoe morele afwegingen in de omgang met AI-systemen aan te brengen zijn. Aan de TU Delft is sinds vorig jaar een compleet centrum met die vraag bezig, het Digital Ethics Centre. Oprichter en directeur is Jeroen van den Hoven, hoogleraar ethiek en technologie aan de TU. Hij zegt: „Iedereen praat nu over algoritmes, maar in feite is dat *too little, too late*. Onze benadering gaat uit van AI-toepassingen als socio-technisch

systeem dat je als geheel moet ontwerpen. Met hardware, softwareregels én de menselijke rol tegelijk.”

Dat centrum groeit als kool, zegt Van den Hoven, nu het wetenschappelijke en publieke denken over AI en ethiek in een stroomversnelling is gekomen. Er werken zo’n tien mensen, die zowel fundamenteel als toegepast onderzoek doen. Tot het laatste horen projecten voor onder meer de uitkeringsinstantie UWV en het ministerie van Buitenlandse Zaken, met onderzoek naar moreel verantwoord gebruik van AI op militair gebied. Het onderstreept de beweging naar *design* in de ethiek en de techniekfilosofie, die zich heeft ontwikkeld van abstracte bespiegelingen over de culturele effecten van technologie naar meer aandacht voor verantwoord ontwerp en de interactie tussen mens en techniek. „Technologische ontwerpen zijn nu eenmaal niet waardeneutraal en onze waarden hebben als het goed is consequenties voor het ontwerp.”

Lees ook:
**Als filosoof van de informatica had Edsger Dijkstra een trouwe groep volgelingen**

Nederland loopt daarin voorop, volgens de Delftse hoogleraar. „Je kunt gerust zeggen dat wij de bakermat zijn van dit soort denken over techniek„ Het is ook heel Nederlands, zegt hij: aandacht voor techniek én ethiek, ook in militaire zaken. Van der Hoven geeft het voorbeeld van het Amerikaanse marineschip dat in 1988 per abuis een Iraans passagierstoestel neerschoot, met 290 doden tot gevolg. Die ramp had kunnen worden voorkomen met een *battle management*-systeem dat „op een kritiek moment relevante informatie geeft als

context voor lastig te interpreteren data". Zoals: een icoontje dat toont of een toestel stabiel horizontaal nadert (en dus mogelijk een burgertoestel is), of scherp verticaal dalend (een aanvallend militair toestel). *Explainable AI* moet gebruikers in staat stellen beter te begrijpen hoe het antwoord van een complex en ondoorzichtig AI-systeem tot stand komt en hoe betrouwbaar het is.

Van den Hoven: „Morele afwegingen moeten al in de ontwerpfase van een systeem worden meegenomen." Ethiek en techniek kunnen ook van nature heel goed samengaan, zegt hij. „Ingenieurs en moraalfilosofen hebben gemeen dat ze zich een werkelijkheid voorstellen die er nog niet is, maar die er zou moeten zijn." Alleen waren ethici traditioneel „vooral bezig met het doel, niet met de weg ernaartoe, zoals ingenieurs. In het werken aan verantwoorde AI gaat dat samen".

Zulke moreel-technische innovatie, echt iets nieuws neerzetten, vindt Van den Hoven interessanter dan louter kauwen op abstracte kluiven zoals de vraag of een computer kan denken of bewustzijn heeft. „Edsger Dijkstra, een beroemd informaticus en winnaar van de Turing Award - de Nobelprijs voor informatici - zei eens dat hij de kwestie of computers kunnen denken zoiets vond als je afvragen of onderzeeboten kunnen zwemmen. We hebben die dingen en ze kúnnen wat. Daar zullen we ons toe moeten verhouden."

> *Morele afwegingen moeten al in de ontwerpfase van een systeem worden meegenomen*
>
> **Jeroen van den Hoven** –hoogleraar ethiek en technologie

## GEMARGINALISEERDE GROEPEN

Dat neemt niet weg dat het Delftse centrum ook fundamenteel onderzoek doet naar epistemologische vraagstukken en meta-ethiek. Een belangrijke rol daarin speelt het jonge morele begrip *epistemic injustice*. Dat slaat op het gevaar, dat ook Broersen noemt, dat kennis en belangen van zwakke of gemarginaliseerde groepen onvoldoende worden meegewogen. Dat is een reëel risico, zeker bij grootschalige taalprogramma's. Van den Hoven kreeg onlangs een testje onder ogen dat ernstige bias blootlegde in ChatGPT's aanpak van het 'trolley-probleem' (de vraag of de koers van een losgeslagen trolley moet worden verlegd zodat hij minder slachtoffers maakt). „Dan blijkt dat Chat GPT neigt naar het niet redden van vrouwen, of mensen met een handicap."

Ook zulke vooringenomenheid moet te voorkomen zijn in de ontwerpfase, zegt Van den Hoven. De les is: articuleer ethische overwegingen bij het ontwerp en de ontwikkeling van AI, niet pas naderhand. Hij hoopt dat dan de tijd voorbij is dat nieuwe technologie alijd „tragische keuzes" met zich meebrengt, zoals tussen economische groei óf duurzaamheid, privacy óf veiligheid. „Dat is de potentie van deze technologie. Met een simpele vergelijking: een paraplu maakt op een regenachtige dag twee doelen tegelijk mogelijk, je kunt gaan

wandelen én je blijft droog. Zónder moet je de ‘tragische’ keus maken: binnen droog blijven of wandelend nat worden.” Hij wil maar zeggen: „Verantwoorde AI ís de paraplu.” Werken met innovatieve technologie hoeft geen „compromis” te zijn of *trade-off* tussen botsende belangen.

Ook verder naar het oosten, in Wageningen, zijn ethici in het laboratorium te vinden. Ze werken er met ingenieurs onder meer aan AI-applicaties voor ‘precisielandbouw’. Via sensoren in melkrobots kunnen gezondheidsdata worden verzameld van individuele koeien in een veehouderij, vertelt filosoof Vincent Blok, onlangs benoemd tot hoogleraar. Met een groep onderzoekers bestudeert hij ook de adaptatie van AI door werknemers in grote bedrijven, wat *epistemic injustice* moet tegengaan. „Je kunt AI-applicaties nooit zomaar invoeren”, zegt Blok, „mensen moeten erbij worden betrokken en inbreng kunnen hebben. Technologische innovatie is een co-creatie.”

> *Mens en machine groeien naar elkaar toe, onmiskenbaar*
>
> **Vincent Blok** –hoogleraar disruptieve technologieën

## EUROPESE TRADITIE

Terwijl Broersen en Van den Hove [illegible] analytisch-filosofische [illegible] en accent op logica en epistemologie [illegible] de Europese continentale traditie [illegible] de toon over techniek en technologie werd daarin lange tijd door speculatieve cultuurfilosofen als Martin Heidegger en Jacques Ellul. Zij interpreteerden moderne techniek niet als instrument voor een vastgesteld doel, maar als een systeem dat zijn eigen werkelijkheid produceert. Of als een vraatzuchtige moloch die de hele menselijke wereld dreigde te koloniseren. Hun invloed is nog volop aanwezig in het werk van denkers als Paul Virilio en Bruno Latour.

Blok plaatst zichzelf genuanceerd in die school. Met zijn onderzoeksgroep wil hij over AI niet vooral denken vanuit de concrete applicaties, maar vanuit een „filosofisch systeemdenken, zonder de artefacten los te laten”. Dat houdt in: AI zien als technologie die een nieuwe werkelijkheid opent. Blok: „Technologie is nooit alleen maar een middel, het opent een nieuwe beleving van de werkelijkheid.” Een goed voorbeeld is de anticonceptiepil: een vondst bedoeld om de bevolkingsgroei af te remmen, die een sociaal-seksuele revolutie teweegbracht. Blok: „Neem zoiets eenvoudigs als een bril. Je gaat niet alleen beter zien, je stapt een nieuwe wereld binnen. Techniek medieert de ervaring.”

Groot verschil met die oudere systeemfilosofen is dat de praktijk van AI de denkers nu steeds meer op de hielen zit. Blok: „Er is door filosofen veel en diep nagedacht over kunstmatige intelligentie. Maar dat was allemaal nog heel abstract. Nu hebben we er de concrete ervaring bij, we kunnen het beleven." Dat kan verbluffen, betoveren én ontnuchteren tegelijk. „We kunnen nu dus ook concluderen: nee, dit is het bij nader inzien toch niet helemaal."

Welke wereld opent AI? In zijn recente boek *Reality +* beweert David Chalmers, een populaire expert in *philosophy of mind*, dat er op de keper beschouwd geen ontologisch verschil is tussen 'echte' en virtuele werkelijkheden. Virtuele mensen zijn in een digitale simulatie even 'echt' als die van vlees en bloed in de gewone wereld. Uiteindelijk is ook die laatste volgens hem opgebouwd uit bits en een soort simulatie. Blok: „Dat is wat pan-computationalisten geloven: alles is data en dus te berekenen." Het illustreert, zegt hij, de *loop* die is opgetreden in het denken over intelligentie. „Toen computers nieuw waren, probeerden we ze te begrijpen vanuit menselijke intelligentie: kunnen ze wat wij kunnen? Inmiddels zie je het omgekeerde gebeuren: we interpreteren menselijke intelligentie nu aan de hand van wat computers doen: dataverwerking."

Hij denkt daar anders over. Ook Blok ziet de grens tussen mens en machine wel vervagen maar niet verdwijnen. „Mens en machine groeien naar elkaar toe, onmiskenbaar. Maar is wat een machine doet vergelijkbaar met onze intelligentie, of hooguit met een uitsnede ervan?" Menselijke intelligentie is vergeleken met het licht van een gloeilamp, die breed om zich heen schijnt, die van AI met een laser: intens, maar nauw.

Lees over filosoof Vincent Blok:
**Moeder natuur is niet lief, ze is keihard en onvoorspelbaar**

## ONEINDIG KRACHTIGER

Mensen zijn „semantische wezens”, zegt Blok, die intuïtief betekenissen zien, intenties herkennen en verbanden leggen die een oneindig krachtiger rekenmachine mist. Blok geeft een ander voorbeeld: liefdesverdriet. „Wij hebben die pijn omdat we wéten wat we verliezen, we kennen de keerzijde. Dat is inherent aan menselijke beslissingen. Een *decision-support system* kent dat niet.”

Blok ziet zo inzichten terug uit het werk van fenomenologen als Edmund Husserl en Merleau-Ponty, die de ‘intentionele’ structuur van de menselijke ‘leefwereld’ in kaart brachten. „Het is leuk om daar met *engineers* over te praten. Softwaremensen snappen meteen wat je bedoelt met een ‘semantisch veld’. Dat is een heel andere benadering dan van buitenaf iets inbouwen.”

Dat keert bij alle drie de filosofen terug: AI vraagt om een integrale benadering’. Blok: „Je zou AI ook ecologisch kunnen zien, niet langer gecentreerd rond de mens, maar biocentrisch. Het gebruik van kunstmatige intelligentie slurpt enorm veel energie, is dat altijd te verantwoorden?” Het zwaartepunt in zulke keuzes kun je ook leggen bij twee menselijke polen: individuele voordelen of het collectieve belang. Blok: „Je moet die twee telkens opnieuw afwegen. Je kunt voorspellende AI gebruiken om stropers in een natuurgebied te

bestrijden, maar ook voor *predictive policing* of om dissidenten op te sluiten. Filosofen kunnen daar veel over nadenken, maar uiteindelijk is het een vraag voor de hele samenleving. Wat willen we wel en wat niet?”

Foto’s **Dieuwertje Bravenboer**

## KUNSTMATIGE INTELLIGENTIE

**Is kunstmatige intelligentie ‘de toekomst’?** Robot ChatGPT maakte duidelijk dat technieken die het menselijk denken overtreffen al veel verder zijn dan wij ons realiseren. Deze serie maakt de balans op: welke rol gaan machine learning en neurale netwerken spelen in de wetenschap? Opnieuw is de conclusie: we zijn al door de toekomst ingehaald.

1. De volgende Einstein kan een computer zijn, maar hoe lang duurt dat nog?
2. Overal langs zijn behandeltraject komt de patiënt kunstmatige intelligentie tegen
3. AI zit verblufte, betoverde én ontnuchterde filosofen op de hielen
4. Uit een berg vage signalen kan AI afleiden dat de aarde gevaarlijk wordt
5. Column Beatrice de Graaf: Het tomeloze technologische vooruitgangsoptimisme
6. Camera’s en computers kunnen de natuur volautomatisch analyseren

*Een versie van dit artikel verscheen ook in de krant van 25 februari 2023*