

Artificiële intelligentie. Experts relativeren onheilspellende toekomstscenario's en zetten blik op scherp

AI heeft zijn gevaren, maar ze schuilen in een andere hoek

Zijn we, nu de ontwikkeling van artificiële intelligentie in een stroomversnelling komt, nog maar een duimbreed verwijderd van machines die slimmer worden dan wij en onze ondergang inluiden? 'Dystopische toekomstscenario's leiden de aandacht af van de échte problemen, die er nu al zijn.'

DIETER DE CLEENE

A

rtificiële intelligentie is "een dringender bedreiging" voor de mensheid dan de klimaatverandering, als we AI-pionier George Hinton mogen geloven. In het slechtste geval is het "licht uit voor ons allemaal", denkt OpenAI-CEO Sam Altman. Volgens de top van dat bedrijf is zelfs een equivalent van het Internationaal Atoomagentschap nodig om de mensheid te beschermen tegen 'superintelligente' AI.

Samen met de capaciteiten van AI neemt ook het aantal apocalyptische waarschuwingen toe. Vorige maand riepen grote namen uit de techwereld op om de ontwikkeling van AI-systemen krachtiger dan GPT-4 zes maanden stil te leggen. Ze wezen op 'ernstige risico's voor de samenleving en de mensheid'. Onder de ondertekenaars: Twitter-, SpaceX- en Tesla-CEO Elon Musk, Apple-oprichter Steve Wozniak en verschillende onderzoekers bij Googles Deepmind.

Musk en co. wijzen erop dat AI-systemen steeds beter met mensen kunnen wedijveren. Moeten we niet-menselijke breinen ontwikkelen die uiteindelijk slimmer kunnen worden dan wij en ons overbodig kunnen maken en vervangen?, vragen ze zich af. Moeten we een totaal controleverlies over onze samenleving riskeren?

Ook historicus Yuval Harari tekende de oproep. De Israëlische bestsellerauteur waarschuwde eind maart in *The New York Times* al dat AI die taal beheerst ons kan opsluiten in een 'wereld van illusies', met alle mogelijke gevolgen van dien. Hij verwijst ook naar een bevraging uit 2022, waarbij de helft van de deelnemende academici en AI-onderzoekers aangaf dat er een kans van minimaal 10 procent was dat AI tot het uitsterven van de mens zou leiden.

Volgens Hinton staat het overleven van de mens op het spel wanneer computers slimmer worden dan wij. "AI zal ons mis-



© STUDIO SW

schien nog even tolereren om de elektriciteitscentrales draaiend te houden”, vertelde de voormalig Google-werknemer aan de *MIT Technology Review*. “Maar na een tijdje misschien niet meer.” Slik?

Bullshit”, reageerde techniekfilosoof Mark Coeckelbergh (Universiteit Wenen) eerder al op de onheilsprofetieën van Harari. Hij stelt vast dat de paniek over superintelligente computers ver van de realiteit staat. “Enkele invloedrijke personen en de media propageren die science-fictionsscenario’s”, zegt Coeckelbergh. “Maar die dystopische toekomstbeelden leiden de aandacht af van de echte problemen, die er nu al zijn.”

KRITI-HYPE

AI-experts Sayash Kapoor en Arvind Narayanan (beiden Princeton University) noemen de open brief van Musk en co. een staaltje ‘kriti-hype’: de brief is kritisch voor de ondoordachte ontwikkeling van AI, maar hypet tegelijk de mogelijkheden ervan, en stelt die als veel spectaculairder voor dan ze werkelijk zijn. De ontwikkelaars van AI lijken zo “scheppers van buitenaardse werktuigen”.

En wie speelt dat in de kaart? Juist. “Die AI-hype helpt diegenen die deze systemen bouwen ze te verkopen”, stelt hoogleraar Computatieve Linguïstiek Emily Bender (University of Washington) in een reactie op de open brief.

Om te beginnen valt er wel wat aan te merken op de waarschuwing voor ‘niet-menselijke breinen’ die onze intelligentie overtreffen. Over wat intelligentie precies inhoudt, kunnen wetenschappers en filosofen oeverloos discussiëren. “In bepaalde toepassingen overtreft AI de mens nu al”, zegt techniekfilosoof Katleen Gabriëls (Universiteit Maastricht). “Maar de menselijke intelligentie is erg veelzijdig en complex.

Wie waarschuwt voor computers die op alle vlakken superieur zijn, overschat waartoe AI in staat is.”

Het gevaar van AI is volgens Coeckelbergh niet zozeer dat systemen té goed zijn, maar dat we verblind door de hype de gebreken van ontransparante systemen niet langer zien. Dat is bijvoorbeeld het geval wanneer mensen een chatbot als ChatGPT als een alwetend orakel gaan zien. Terwijl zo’n taalmodel enkel is getraind om plausibel klinkende tekst uit te braken, niet om de waarheid te vertellen.

Het kan ook problematisch zijn als algoritmes worden ingezet om bij sollicitatieprocedures de meest geschikte kandidaat te selecteren, of om te bepalen wie in aanmerking komt voor een lening. De Britse krant *The Guardian* meldde onlangs dat een derde van de Australische bedrijven AI gebruikt bij aanwervingen. “Dat soort systemen is in opmars”, zegt Coeckelbergh. “Het risico op discriminatie loert dan om de hoek.”

Amazon voerde eerder al een systeem om kandidaten te rangschikken af. Omdat het was getraind met de cv’s van voornamelijk mannelijke werknemers uit de techsector, benadeelde het systematisch

‘We moeten de ethische problemen aanpakken. Dat is een groter gevaar dan superintelligente computers’

MARK COECKELBERGH
UNIVERSITEIT WENEN

vrouwen. “Zo’n ontransparant model kan een rookgordijn zijn waar een bedrijf zich achter kan verstoppen”, zegt computerwetenschapper Jeroen Baert. Zo kan AI u een job kosten, maar dan niet omdat het ‘alle jobs weg automatiseert’, zoals de ondertekenaars van de open brief waarschuwen.

Dat artificiële intelligentie een ingrijpende impact zal hebben op jobs staat buiten kijf. Sommige zullen verdwijnen of op zijn minst ingrijpend veranderen, maar er zullen ook nieuwe bij komen. Maar ook in dit geval mogen onheilsplende toekomstscenario’s volgens critici de aandacht niet afleiden van de manieren waarop AI nu al banen bedreigt.

De Canadese schrijfster en activiste Naomi Klein klaagt in *The Guardian* aan hoe de techbedrijven hun systemen trainen met alle online beschikbare tekst en beeld, zonder zich veel aan te trekken van eigendomsrecht. “De grootste diefstal in de menselijke geschiedenis, verpakt als een geschenk”, aldus Klein.

In een open brief uitten artiesten eerder deze maand in het bijzonder hun ongenoegen over AI-systemen die beelden genereren, na te zijn getraind met miljoenen afbeeldingen waarop auteursrecht rust. “Terwijl het aantal illustratoren dreigt te decimeren, verdienen de bedrijven achter die AI-systemen fortuinen”, klagen de ondertekenaars die ‘kunstroof’ aan.

En terwijl de AI-pioniers orakelen over toekomstige gevolgen voor onze jobs, zwijgen ze zedig over de onderbetaalde werknemers in de zogenoemde annotatiefirma’s. Hun job bestaat erin voorbeelden van haatspraak, geweld en seksueel misbruik als dusdanig te labelen, zodat een systeem als ChatGPT leert die te herkennen en weg te filteren. Het magazine *Time* onthulde eerder dit jaar hoe Keniaanse werknemers die in onderaanneming voor OpenAI werkten daar minder dan twee dollar per uur voor kregen. Achter de blinkende AI-façade zit soms psychologisch erg belastende uitbuiting, aldus *Time*.

Nog zo’n onderbelicht probleem: de milieupact van AI. Want de training en het gebruik van AI-systemen met steeds meer rekenkracht en grotere datasets vreet veel energie. De training van ChatGPT3 zou volgens wetenschappers aan de University of California zo’n 550 ton CO₂ hebben uitgestoten, equivalent aan ruim 200 retourvluchten Brussel-New York. “De groeiende rekenkracht van AI is reden tot bezorgdheid om de duurzaamheid ervan”, klinkt het in een recent rapport van het Global Partnership on Artificial Intelligence.

MENS BUITEN SCHOT

De mogelijkheden van AI opblazen leidt er niet alleen toe dat mensen kritiekloos vertrouwen op wat een systeem zoals ChatGPT hen voorschotelt en dat ze de indruk krijgen dat er een denkend wezen achter schuilgaat, waarschuwen Bender en collega’s aan het Distributed AI Research Institute. Zo’n discours legt de verantwoordelijkheid ook te veel bij de AI-systemen zelf, en te weinig bij hun bouwers.

Baert merkt op dat er aan een scenario waarin AI de wereld om zeep helpt een heleboel menselijke beslissingen voorafgaan. “Mensen bouwen die systemen,

trainen ze en beslissen om ze op een bepaalde manier in te zetten”, zegt Baert. “Doembeelden minimaliseren die menselijke rol.”

Musk en co. waarschuwen voor ‘de dramatische economische en politieke verschuivingen die AI zal veroorzaken’. Maar als dat zo is, zal dat niet te wijten zijn aan AI, vindt Bender. Wel aan de bedrijven die AI op een zo winstgevend mogelijke manier willen inzetten. Als ChatGPT voor problemen zorgt en desinformatie spuit, ligt dat niet aan het systeem zelf, merkt ze fijntjes op. Wel aan de makers die hebben besloten het als makkelijk toegankelijke tool op de wereld los te laten.

Dat is ook wat AI-expert Johan Loeckx (VUB) stoort. “De techbazen framen het debat zo dat ze zelf buiten schot blijven”, zegt Loeckx, verbonden aan het Artificial Intelligence Lab. “Alsof de AI-systemen zelfstandig handelende entiteiten zijn. Terwijl wij bepalen hoeveel autonomie en controle we uit handen geven.”

Dat de sector zelf oproept tot regulering lijkt sympathiek. Maar zo nobel zijn de

‘Mensen bouwen die AI-systemen en trainen ze. Doembeelden minimaliseren die menselijke rol’

JEROEN BAERT
COMPUTERWETENSCHAPPER

motieven volgens de Britse hoogleraar John Naughton niet. Als CEO’s om regulering vragen, is dat doorgaans een openlijke sollicitatie om zelf een belangrijke rol te spelen in de totstandkoming van die wetgeving, merkt hij op in *The Guardian*.

En kijk. Voormalig Google-CEO Eric Schmidt liet zich op de Amerikaanse televisiezender NBC alvast ontvallen dat “iemand van buiten de industrie onmogelijk kan begrijpen wat er allemaal mogelijk is”. “Lariekoek”, vindt Baert. “Als dat zo is, dan is het aan de industrie om beter te communiceren, en meer informatie te geven over hoe systemen werken en getraind zijn.”

De techsector blinkt sowieso al niet uit in transparantie over hoe ze algoritmes op platformen zoals Facebook en YouTube inzet om gebruikers aan het scherm gekluisterd te houden, en wat ze daaruit over ons leert. Diezelfde bedrijven krijgen met steeds betere AI nu een nog krachtiger wapen in handen. “Dat kan de macht van Big Tech nog vergroten”, vreest Gabriëls.

AI-systemen zijn vaak goed in verbanden leggen tussen verschillende grote datasets. Zo kunnen ze bijvoorbeeld worden gebruikt om geanonimiseerde data te ‘de-anonimiseren’ en het gedrag van gebruikers te voorspellen. “Een bedrijf zoals Google weet nu al veel over ons”, zegt Loeckx. “In essentie kan je met AI meer over iemand te weten komen, op basis van minder data. Dat brengt uiteraard privacyrisico’s met zich mee.”

Helemaal problematisch wordt het als dubieuze regimes AI inzetten om burgers te controleren en te manipuleren. “AI biedt de perfecte tools om een surveillancestaat uit te bouwen”, zegt Loeckx.

Dat AI ook veel moois te bieden heeft, van betere geneeskunde tot het verlichten van saaie arbeid, staat buiten kijf. Dat er een debat en regelgeving moeten komen om de gevaren te beperken ook. Belangrijk is volgens Coeckelbergh om daarbij de juiste klemtoon te leggen. “We moeten de concrete ethische problemen aanpakken die veel mensen kunnen treffen. Dat is een groter gevaar dan superintelligente computers die ontsporen en de mensheid vernietigen.”

