



De AI-hype dwingt tot nadenken over
ethisch en mensgericht innoveren

WAS CHATGPT EEN GOED IDEE?

Slimme chatbots, zelfrijdende wagens, digital twins: in de AI-sector volgen de innovaties elkaar aan een duizelingwekkend tempo op. Maar wat technologisch mogelijk is, is maatschappelijk niet altijd wenselijk. Om risicovolle ontwikkelingen te voorkomen, wil Europa regels opleggen aan AI-ontwikkelaars. Maar hoe vind je de balans tussen innoveren en reguleren?

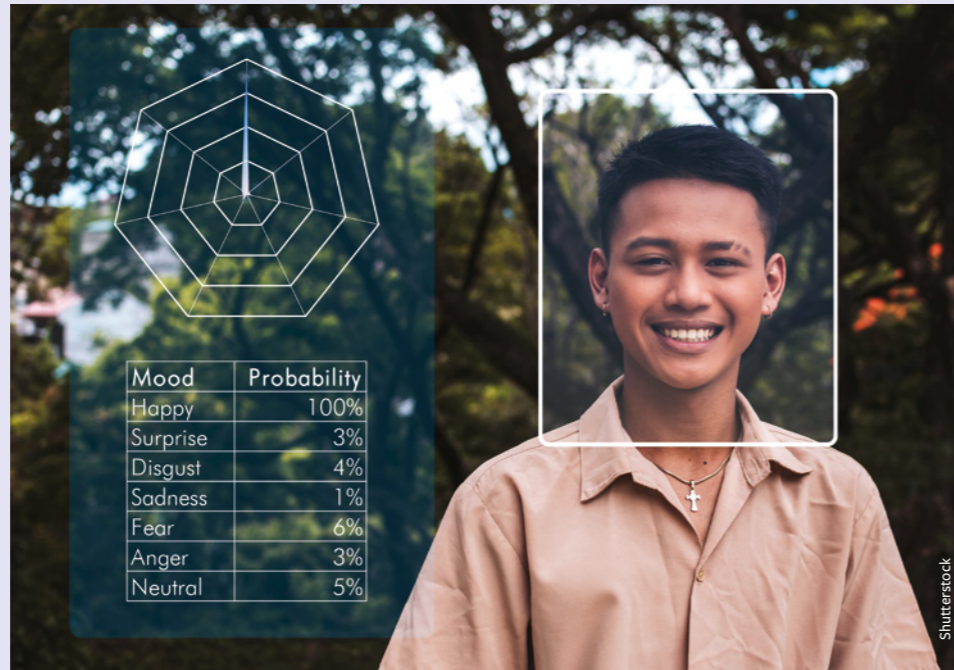
Peter Vanwijnsberghe

Welke uitvinder heeft eigenhandig de meeste schade berokkend aan het milieu? Die bedenkelijke eer valt wellicht Thomas Midgley Jr. ten deel. De Amerikaanse ingenieur en scheikundige ontdekte iets meer dan honderd jaar geleden dat tetraethyllood de klopvastheid van een automotor verhoogt. De auto-industrie was in de wolken. Vanaf 1923 begon ze het loodmengsel op grote schaal als additief toe te voegen aan benzine. Decennialang kwam de giftige stof via de uitlaatgassen van miljoenen auto's in het milieu terecht.

Enkele jaren later stond Midgley aan de wieg van chloorfluorkoolwaterstoffen (cfk's). Dat gas moest een veilig alternatief vormen voor de giftige koudemiddelen die tot dan in koelkasten werden gebruikt. Pas in 1974 – Midgley was al dertig jaar overleden – ontdekten wetenschappers welke desastreuze effecten cfk's hebben op de ozonlaag.

Midgley was een briljant uitvinder: met zijn werk ontketende hij een revolutie in zowel de auto- als de chemische industrie. Hij bedacht doeltreffende oplossingen voor bestaande problemen. Maar op de gevolgen voor mens en milieu had hij zich niet bezonnen.

Vandaag schittert een nieuwe ster aan het innovatiefirmament, en ze lijkt helderder te schijnen dan alle tetraethyl en cfk's samen: artificiële intelligentie. Op korte tijd vuurden AI-systemen in talloze domeinen innovaties aan. Zelfrijdende wagens moeten jaarlijks talloze verkeersongevallen voorkomen. Met AlphaFold kunnen we de 3D-structuur van miljoenen eiwitten accuraat voorspellen. Digitale simulatiemodellen van de planeet helpen de klimaatverandering in kaart te brengen en gericht



actie te ondernemen. Het potentieel van AI lijkt wel bodemloos te zijn.

Tegelijk komen de doemscenari'o's steeds scherper in beeld. Miljoenen *microworkers* trainen software voor zelfrijdende auto's, transcriberen gesprekken of labels afbeelden. Ze zijn systematisch onderbetaald en ondervetegenwoordigd. De servers waarop AI-systemen draaien, slurpen water en vreten elektriciteit. Onlangs rekenden onderzoekers voor dat ze tegen 2027 jaarlijks naar schatting zowat evenveel stroom verbruiken als heel Nederland. En dan zijn er nog de drie D's: desinformatie, datamisbruik en deepfakes. De impact van artificiële intelligentie op de samenleving is gigantisch. Hebben we ons met deze technologie een nieuw nuttig gevaar op de hals gehaald, net als met tetraethyl en cfk's? En zo ja: zijn we ons daar voldoende van bewust?

'We bevinden ons op een kantelpunt', zegt technologiefilosoof Mark Coeckelbergh (Universiteit Wenen), wiens boek *AI Ethics* in 2022 verscheen. Artificiële intelligentie neemt een hoge vlucht. Hoe we vandaag met die ontwikkelingen omgaan zal bepalend zijn voor de komende

periode, denkt hij. 'Er zijn bijzonder veel kansen en mogelijkheden. Generatieve AI-modellen als ChatGPT lijken daar eindelijk ook het grote publiek van te hebben overtuigd', zegt Coeckelbergh. 'Maar big tech heeft momenteel wel heel veel macht. Veel experts maken zich zorgen, en niet het minst over die slimme chatbots. Hoeveel jobs staan hier op de helling? Hoe beschermen we ons tegen grootschalige manipulatie? Wie zullen de winnaars en verliezers zijn van deze transformatie?'

Die laatste vraag is bijzonder relevant, vindt ook jurist en filosoof Nathalie Smuha (KU Leuven en New York University). 'Heel wat betrokkenen benadrukken de voordelen van de technologie. Die zijn er ook, ten overvloede. Maar we moeten ons wel afvragen: wie definieert wat een voordeel is, en wie kan ervan profiteren? Niet iedereen plukt de vruchten van innovatieve ontwikkelingen, sommige individuen of groepen blijven in de kou staan, of zijn regelrecht benadeeld.'

Een vrij recente ontwikkeling die Smuha zorgen baart, is emotieherkenning, een techniek waarmee emoties of gemoedstoestanden van

Met behulp van emotieherkenning kan een AI-model nagaan of iemand liegt, of zelfs welke geaardheid die heeft. Deze ontwikkeling zou sterk gereguleerd worden in Europa.

het gezicht worden afgelezen. Het AI-model kan een geopende mond, opkrullende mondhoeken en rimpels rond de ogen registreren als kenmerken van lachen, waaruit het model vervolgens opmaakt dat de geanalyseerde persoon blij is.

Via zo'n tool kan een blinde of slechtziende misschien meer vat krijgen op zijn of haar omgeving, zegt een deskundige in het vakblad *MIT Technology Review*. Maar ze kan ook worden gebruikt om na te gaan of iemand liegt, of zelfs welke geaardheid die heeft. 'Zulke zaken kun je niet zomaar van iemands gezicht aflezen, dat is niet wetenschappelijk onderbouwd', zegt Smuha. 'Dat het toch wordt gedaan is problematisch, zeker als je weet hoe sommige politieke leiders aankijken tegen bijvoorbeeld homoseksualiteit.'

Grip op de zaak

Ongebreidelde innovatie in de AI-sector is niet zonder risico's. Van die boodschap zijn ook steeds meer beleidsmakers doordrongen. Eind 2023 bereikten het Europees Parlement en de Raad van de EU een principiële akkoord over de AI Act. Die nieuwe wet moet de lidstaten helpen om de voordelen van de technologie te benutten en moet hen tegelijk tegen de mogelijke risico's beschermen. De wetgeving zou in 2026 in werking treden.

'Met de AI Act zegt Europa: we willen innovatie stimuleren, maar

niet tot elke prijs', aldus Smuha, die als coördinator van een deskundigengroep de richtlijnen die de AI Act inspireerde mee vormgaf. 'Als AI-systemen inbreuk dreigen te plegen op onze basisrechten, op democratische waarden, de rechtsstaat of het milieu, dan moeten die worden gereguleerd of verboden.'

De emotieherkenning waarover Smuha het hierboven had, zou sterk worden gereguleerd in Europa. Net als systemen die op basis van hun gedrag 'sociale scores' toekennen aan individuen of groepen. Voor deze en andere ontwikkelingen voorziet de AI Act in risicocategorisering. Systemen die een zeker gevaar met zich meedragen voor fundamentele waarden moeten bijvoorbeeld een verplichte impactanalyse ondergaan voordat ze op de markt kunnen komen. Aan ontwikkelingen met een lager risico wordt enkel transparantie opgelegd. Voor systemen die geen enkel risico stellen, zijn er geen regels.

Als ChatGPT onder de AI Act zou verschijnen, zou het vooraf aan grondige evaluaties worden onderworpen. Terecht, vindt Smuha. 'Ik weet niet in hoeverre de samenleving vandaag echt gebaat is met generatieve AI. We hebben nog te weinig zicht op de negatieve gevolgen, zeker op langere termijn. Dat komt deels doordat ontwikkelaars als OpenAI niet eerst een risicoanalyse uitvoerden. In de toekomst moeten we ontwikkelaars verplichten hier meer verantwoor-

delijkheid voor te nemen.' Dat zegt ook Coeckelbergh: 'Eigenlijk zouden bedrijven al bij het ontwikkelen ethische waarden in overweging moeten nemen, nog voor er van regulatie sprake is.'

Dat Europa met de AI Act kritisch nadenkt over welke innovaties het wel en niet wil, vindt Smuha een goede zaak. 'Niet elke innovatie is per definitie een verbetering. Een wetgevend kader met een doordachte ethische basis helpt om artificiële intelligentie de richting uit te sturen die we als samenleving op willen.' De AI Act kan grip op de zaak bieden.

Dat gevoel wordt vandaag node gemist. Filmanimators maken zich zorgen over de opmars van text-to-imagemodellen, ouders en leerkrachten vragen zich af of kinderen wel zelf hun huiswerk maken, alleenstaande zeventigers hopen dat ze in de supermarkt of in het postkantoor door een mens worden geholpen. Heel wat mensen zien AI-ontwikkelingen tegemoet met een mengeling van onzekerheid, angst en wantrouwen.

Misschien biedt wetgeving dan een zekere geruststelling. 'Als we bepaalde vereisten stellen aan AI-systemen, dan kunnen we de producten en diensten die daar effectiever aan beantwoorden ook makkelijker vertrouwen', zegt Coeckelbergh. 'Dat is dan weer positief voor de bedrijven die ze ontwikkelen: tools en services die breed worden aanvaard hebben ook economisch een betere positie. Europa zet hier een stap in de goede richting.'

'Die stap kan een voorbeeld zijn voor andere landen', vult Smuha aan. 'Toen de GDPR-wetgeving werd doorgevoerd, botste die op veel weerstand. Maar intussen is de GDPR wel een wereldwijd exportproduct geworden. Het zou mooi

'Er gaat nog te weinig aandacht naar de impact van AI op de omgeving en het milieu, met name in niet-westerse landen'

DAVID TEGEN GOLIATH GOOGLE

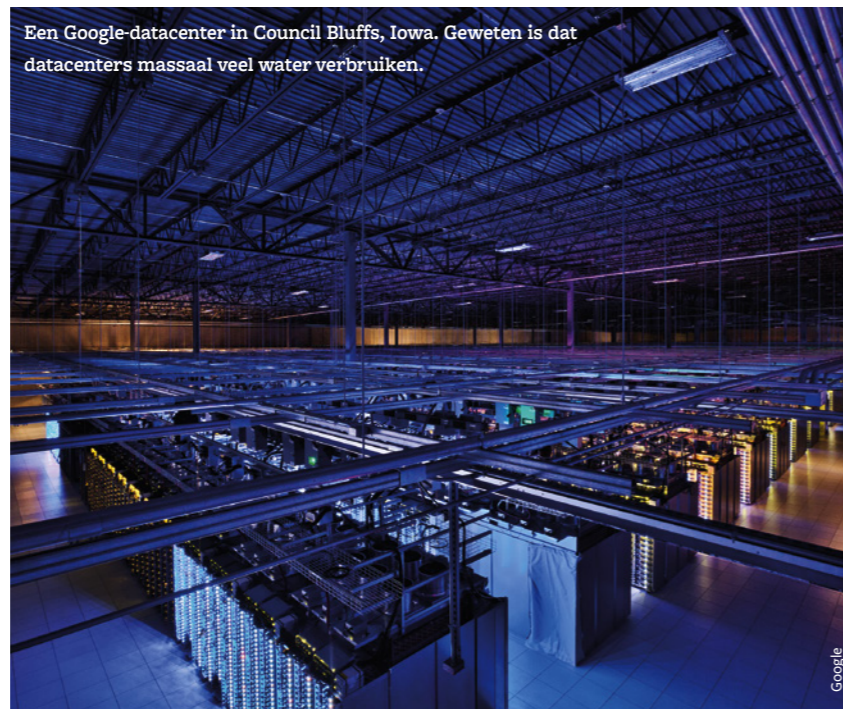
169 liter per seconde. Sebastián Lehuédé was er zelf ook van geschrokken toen hij in 2019 hoorde hoeveel water een nieuw datacenter van Google in Santiago zou verbruiken. Lehuédé, die als docent ethiek, AI en samenleving verbonden is aan het King's College in Londen, was voor andere onderzoeksdoeleinden in de Chileense hoofdstad, maar toen hij via inwoners van de arbeidersgemeente Cerrillos hoorde van de plannen voor het datacenter, begon hij er zich meteen voor te interesseren. Datacenters verbruiken massaal veel water. De servers waarop de data worden verwerkt van onder meer zoekopdrachten, spraakcommando's, navigatie-instructies en trainingen voor machine learning-modellen moeten continu koel blijven. Een manier om dat voor elkaar te krijgen, is via een circulatiesysteem waarin koud water langs de rekken met GPU's, CPU's en geheugenkaarten wordt gepompt en het opgewarmde water naar bijvoorbeeld een koeltoren wordt afgevoerd.

In een vorig jaar verschenen paper rekenen onderzoekers voor dat datacenters gebruikt voor AI-doeleinden tegen 2027 wellicht zo'n 4,2 tot 6 miljard kubieke meter water verbruiken. Dat is de helft van het totale waterverbruik in het Verenigd Koninkrijk.

'Cerrillos kampt al jaren met watertekorten,' zegt Lehuédé, 'maar daar leek Google weinig rekening mee te houden bij de allocatie voor haar datacenter. De lage vestigingskosten waren een veel belangrijker argument.' Sebastián Piñera, de toenmalige president van Chili, steunde het project, en ook vanuit Cerrillos zelf kwam weinig protest. Met uitzondering van een kleine groep activisten, die zichzelf de Moscat-groep noemde. 'Ze schuimden straatmarkten af, hingen posters op en hielden workshops om de gemeenschap te informeren.'

Er kwam een lokaal referendum – niet naar aanleiding van het datacenter, maar de activisten konden er wel een vraag in smokkelen. Van de respondenten stemde 49 procent tegen de bouw van het datacenter. Hoewel het referendum niet bindend was, leek het toch een effect te hebben gesorteerd. 'Google heeft nooit officieel gecommuniceerd dat het zijn plannen voor het datacenter in Cerrillos zou laten varen, maar tot vandaag is er nog geen steen gelegd', zegt Lehuédé. De activisten lijken hun slag te hebben thuisgehaald. Wat zich afspeelde in Cerrillos is geen geïsoleerd geval. Google, Microsoft en andere technologiebedrijven hopen evengoed elders in Latijns-Amerika datacenters te kunnen neerpoten. Bovendien willen ze ook de lithiumvoorraden in onder meer de Chileense Atacamaoestijn exploiteren. Lithium is een belangrijk batterijonderdeel in elektrische wagens, smartphones en andere slimme

toestellen. Ook bij het winnen van dat mineraal wordt veel water gebruikt, zegt Lehuédé. 'In Zuid-Amerika zien veel inheemse gemeenschappen hoe het bestaan van flamingokolonies en verschillende boomsoorten in gevaar komt doordat bedrijven lithiumvoorraden aanboren.' De impact van de AI-hype op lokale gemeenschappen en hun leefomgeving in niet-westerse landen wordt volgens Lehuédé te weinig opgemerkt. 'Met de AI Act had dat kunnen veranderen. Maar bij zijn categorisering van systemen met een hoog risico lijkt de wet geen rekening te houden met de milieu-impact. Het is cruciaal dat we daar strenger in worden, want vandaag hanteren grote bedrijven op dat gebied een zwijgcultuur. En de stem van de gemeenschappen die daar nadeel van ondervinden wordt niet gehoord. In beleids gesprekken over artificiële intelligentie zouden zij een plek aan de tafel moeten krijgen.'



Een Google-datacenter in Council Bluffs, Iowa. Geweten is dat datacenters massaal veel water verbruiken.

Google

zijn mochten de VS en China akte nemen van het Europese AI-beleid. Ze hebben in elk geval een economische drijfveer: met zo'n 450 miljoen inwoners is Europa een serieuze afzetmarkt voor hun AI-producten.'

Van optimisme naar realisme

Europa is niet de enige speler die pusht voor een meer ethische en mensgerichte houding ten opzichte van artificiële intelligentie. Eind oktober 2023 tekenden Argentinië, Brazilië, Chili, Mexico, Suriname, Venezuela en nog veertien andere Latijns-Amerikaanse en Caribische landen een overeenkomst over AI-richtlijnen. De landen pleiten voor een beleid dat garanties biedt op het vlak van gendergelijkheid, toegankelijkheid, privacy en andere waarden.

Eerder die maand lanceerde China tijdens het derde Belt and Road Forum voor Internationale Samenwerking in Peking een *Global AI Governance Initiative*. In een eerste missiestatement staat onder meer dat respect en gelijkheid gidsende principes moeten zijn bij het ontwikkelen van AI-systemen, en dat er mechanismen moeten komen waarmee AI-tools en -systemen ter verantwoording kunnen worden geroepen. China geeft verder te kennen dat ook niet-westerse en ontwikkelingslanden inspraak moeten krijgen bij de regulering van artificiële intelligentie.

En enkele maanden daarvoor kwamen de zogeheten BRICS-landen (Brazilië, Rusland, India, China en Zuid-Afrika) ertoe overeen een studiegroep op te richten. Die moet nagaan hoe de landen samenstandaarden kunnen opstellen om AI-ontwikkelingen veiliger, betrouwbaarder, controlebaarder en rechtvaardiger te maken.

Sebastián Lehuédé (King's College Londen) ziet die initiatieven

'AI bij het ontwikkelen zouden bedrijven ethische waarden in overweging moeten nemen, nog voor er van regels sprake is'

graag verschijnen. In Chili leerde hij welke impact AI-ontwikkelingen kunnen hebben op lokale gemeenschappen en hun leefomgeving (*zie kaderstuk*). Die blijft vooralsnog onderbelicht.

'Ik bestudeer onder meer de milieuproblematiek rond de datacenters en grondstoffen die voor artificiële intelligentie worden gebruikt. Daarover vind ik in de Europese AI Act nog niet zoveel terug. Techno-optimisme maakt tegenwoordig meer en meer plaats voor realisme: er worden best wel wat gevaren en risico's geschetst rond AI. Dat is een goede ontwikkeling. Maar vaak gaat nog te weinig aandacht naar de impact op de omgeving en het milieu, en dan met name in niet-westerse landen.'

Rem of stimulans?

Ethisch innoveren met oog voor mens en milieu, waar dan ook ter wereld: dat is waar we naartoe willen met artificiële intelligentie. En daarvoor is regulatie nodig, meent alvast Europa. Maar wordt die visie wel gedeeld door AI-bedrijven en -ontwikkelaars? Die kunnen de nieuwe wetgeving evengoed zien als een log mechanisme dat een rem zet op innovatie.

'Ik hoor mensen zeggen dat we onszelf met deze wetgeving in de voet schieten. Dat we in Europa niet meer zullen innoveren en volledig afhankelijk zullen worden van andermans technologie', zegt Nathalie Smuha. 'In zijn huidige vorm is de AI

Act inderdaad een complex beest. De tekst is lang en niet bepaald duidelijk, zelfs niet voor getrainde juristen. Het wordt geen sinecure om ze toe te passen. Zeker voor kleinere bedrijven wordt dat een taaie opdracht.'

'Als de AI Act nieuwe ontwikkelingen in een bepaald domein als hoogrisicosystemen aanmerkt, zie ik het voorvallen dat bedrijven minder geneigd zijn om daarin te investeren. Dan krijg je automatisch minder innovatie op dat terrein', vult Mark Coeckelbergh aan. AI ziet hij het zelf zo'n vaart niet lopen. Hij noemt de zogeheten *regulatory sandboxes* die Europa inbouwt: tools waarmee bedrijven kunnen experimenteren met nieuwe producten onder supervisie van een regulerend orgaan.

Maar: wetgeving kun je omzeilen. Hoe strikter ze is, hoe sneller een onderneming of ontwikkelaar daar mogelijk toe verleid wordt. 'We moeten niet naïef zijn en denken dat regelgeving alle uitwassen zal tegenhouden. Ook met de AI Act zullen ongeoorloofde systemen op de markt komen en gebruikt worden', zegt Smuha. 'Dat is geen reden om geen regelgeving aan te nemen.'

'Dan nog vermoed ik dat er bij de meeste bedrijven wel een draagvlak is voor regulering. Als je als ontwikkelaar onzeker bent of de samenleving wel openstaat voor een nieuw product, dan neem je het risico niet om erin te investeren. In die zin kan juridische zekerheid net stimuleren. Regelgeving en innovatie zijn dan niet tegenstrijdig.' ■