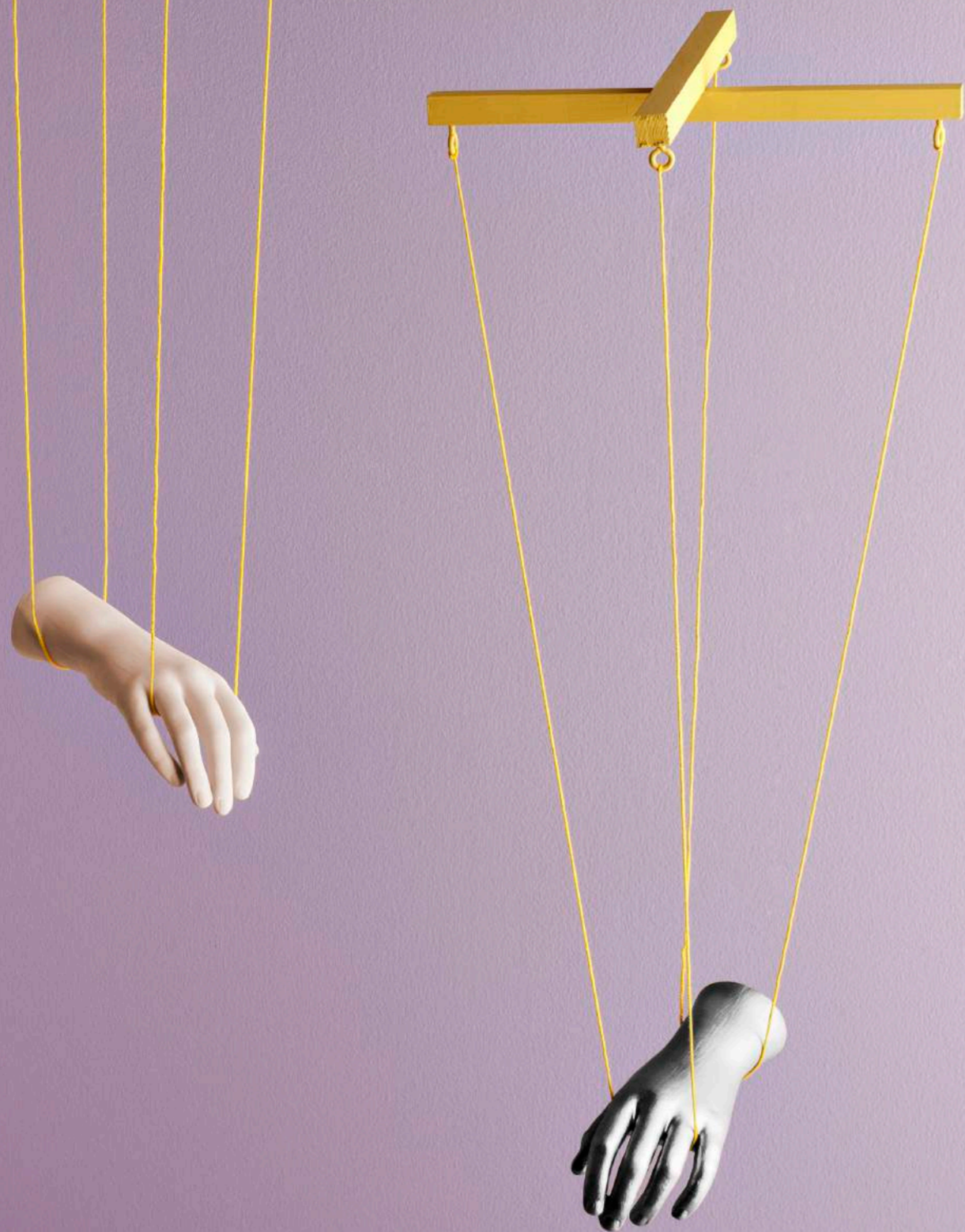


Waar



'Technologie is niet neutraal, maar gevoelig voor bias'

blijft de mens?

Een computer die literatuur doorzoekt, hypothesen opstelt en een medisch behandelplan maakt, noem je dat een collega of een indringer die de mens buitenspel zet? Wat als de computer straks alles beter kan?

'Ik vind het best spannend', zegt Peter Bosman, AI-onderzoeker van het Centrum Wiskunde & Informatica aan de telefoon. 'Komt er een moment dat we té veel op AI gaan vertrouwen?' Dit voorjaar wordt in het Amsterdam UMC de eerste patiënt met prostaatkanker bestraald volgens een behandelplan dat met behulp van kunstmatige intelligentie is uitgerekend. De

computer voert het rekenwerk in een minuutje uit, terwijl dat voorheen een uur in beslag nam. De radiotherapeut, die eindverantwoordelijk is, heeft hierdoor meer tijd om na te denken over de verschillende behandelplannen die de computer voorstelt. Bosman: 'Ik verander mijn werk op een positieve manier: hij krijgt meer inhoudelijke beslissingstijd.' De patiënt heeft mogelijk minder bijwerkingen door een preciezere, patiënt-specifieke bestraling.

Mens blijft eindverantwoordelijk

Tot nu toe kon Bosmans AI nog oefenen met gegevens van reeds behandelde patiënten. Daaruit bleek dat in 98 procent van de gevallen de artsen de voorkeur gaven aan het AI-plan in plaats van een plan dat ze eerder zelf hadden gemaakt. Maar nu de AI 'live' gaat, ontstaat dan niet het risico dat artsen er op termijn volledig op gaan vertrouwen? Denken ze zelf nog wel na? Zoals we gewend zijn geraakt de navigatie in onze telefoon blindelings te volgen, gaat een arts op den duur wellicht ook gedachteloos de uitkomsten van de computer overnemen. Bosman: 'We willen daarom altijd benadrukken dat ons systeem beslissingsondersteunend is. De mens blijft eindverantwoordelijk.' In een notendop laat dit voorbeeld het morele dilemma zien van AI: neemt zo'n systeem niet sluipenderwijs te veel taken van ons over? Hoe behouden we onze vrijheid, creativiteit en autonomie als een computer alles beter kan? En: is het eigenlijk wel erg om taken uit handen te geven?

Wetenschappelijk onzinnig

Volgens Frank van Harmelen, hoogleraar AI aan de Vrije Universiteit Amsterdam, moeten we echt af van het beeld dat machines ons gaan vervangen. 'Dat is maatschappelijk onwenselijk', zegt hij. 'Bovendien is het wetenschappelijk

onzin, want kunstmatige intelligentie is niet te vergelijken met menselijke intelligentie, dus waarom zou je het ene door het andere vervangen?' Dat neemt niet weg dat er wel degelijk een automatiseringsbeweging gaande is, waarin computers eenvoudige taken op zich nemen. 'Maar dat is het gebruikelijke automatiseringsverhaal van de afgelopen tweehonderd jaar. Dat betekent nog niet dat we bestuurd zullen worden door machines.'

Computer als co-auteur

Van Harmelen ziet een alternatieve ontwikkeling voor zich, waarin computers samenwerken met mensen, als een soort collega's. Samen kunnen die beter werk leveren dan wanneer je ze apart aan het werk zet. Bijvoorbeeld in de zorg, de wetenschap en het onderwijs zou zo'n collega goed van pas komen, denkt hij. 'Een computer is in staat om enorme bergen literatuur met elkaar in verband te brengen. Daarin kan hij bijvoorbeeld ontdekken dat in twee totaal verschillende ziektes dezelfde eiwitten een rol spelen. Dat kan leiden tot een alternatieve toepassing voor een medicijn, zoals Viagra ooit voor de bloeddruk werd ontwikkeld maar nu ook voor erectiestoornissen wordt gebruikt.' Om echt goed samen te werken met een mens, moet zo'n computer een *theory of mind* hebben: hij moet het kennisniveau van zijn menselijke 'collega' goed kunnen interpreteren. De komende tien jaar gaan Van Harmelen en vele collega's binnen het NWO-Zwaartekrachtprogramma Hybrid Intelligence onderzoeken hoe mens en machine elkaar op deze manier kunnen versterken in plaats van vervangen. Het doel: over enkele jaren een computer als co-auteur op een wetenschappelijke publicatie.

Nerd lost probleem op

Maar hoe behoudt de mens zijn autonomie naast zo'n snelle, veelzijdige en accurate collega? Bosman denkt dat AI-onderzoekers de belangrijke taak hebben om te benadrukken dat de mens altijd aan de knoppen zit: blijven nadenken en de eindbeslissing nemen. 'Maar hoe je dat bewustzijn ook op de lange termijn vasthoudt, weet ik eigenlijk niet.' Hij merkt dat deze ethische vragen relevant worden, nu zijn AI daadwerkelijk in de klinische praktijk gebruikt gaat worden. 'Wij zijn eigenlijk gewoon

een stel nerds dat een probleem heeft opgelost voor radiotherapeuten, maar we moeten nooit vergeten om onszelf elke keer af te vragen: moeten we dit probleem wel oplossen?'

Privacy, autonomie en rechtvaardigheid

Volgens Fleur Jongepier, universitair docent Digitale Ethiek aan de Radboud Universiteit, is de kans groot dat er toch een soort luiheid optreedt naarmate je een computersysteem langer en vaker gebruikt. Het wordt dan erg verleidelijk om gedachteloos op 'ok' te klikken. Jongepier doet onderzoek naar zelfkennis en autonomie van mensen in de digitale wereld. Volgens haar is het mogelijk om menselijke tussenkomst in het ontwerp van een nieuw systeem te integreren. 'Bijvoorbeeld met een pop-up, die je dwingt om te vertragen en bewust na te denken. Maar het moet ook niet tot te veel frustratie leiden, zoals bij cookie-meldingen op websites. Hier ligt dus nog een leuke uitdaging voor ontwerpers en ingenieurs.' Jongepier merkt dat zij en haar collega-ethici de laatste jaren veel gevraagd worden om mee te denken over nieuwe technologieën. 'Ethici worden inmiddels weggekaapt door bedrijven – soms oprecht, maar ook als een vorm van *ethics washing*. Dan worden ze zonder enige zeggenschap uitgenodigd, of om juridische regulering te voorkomen.' *Ethics first* is dan ook haar adagium. Het ontwerp van een nieuwe technologie moet op waarden gebaseerd zijn, die je eerst met een brede groep belanghebbenden moet formuleren. 'Denk aan privacy, autonomie, solidariteit en rechtvaardigheid. Efficiëntie moet zeker niet altijd het enige uitgangspunt zijn.'

Twee andersoortige computers

Jongepier is allergisch voor zowel de technoutopisten als -pessimisten. 'AI kan ontzettend succesvol zijn en wordt op den duur misschien zelfs een recht. Het is niet ondenkbaar dat patiënten binnenkort een medische behandeling of advies op basis van AI zullen eisen.' Toch ziet ze wel risico's van een AI-systeem als wetenschappelijk co-auteur. 'Technologie is niet neutraal, maar gevoelig voor *bias* ofwel vooringenomenheid. Een AI-systeem zou misschien suggereren om alleen nog maar onderzoek te doen dat binnen twee jaar succesvol uitgevoerd kan zijn, wat ten koste zou gaan van fundamenteel onderzoek.' Van Harmelen is het daarmee eens. 'Hier heb ik ook nog geen oplossing voor. Morele waarden verantwoordelijkheid en uitlegbaarheid moeten vanaf het begin in een ontwikkelingsproject worden meegenomen.' Ondanks dat AI voor de nodige opschudding zorgt in de maatschappij, zijn de menselijke creativiteit en vindingrijkheid beslist niet in het geding, besluit Van Harmelen. 'Denken is uiteindelijk ook gewoon rekenen. Maar mens en machine zijn wel twee heel andersoortige computers, met andere vormen van creativiteit. Die twee kunnen elkaar juist goed aanvullen.'