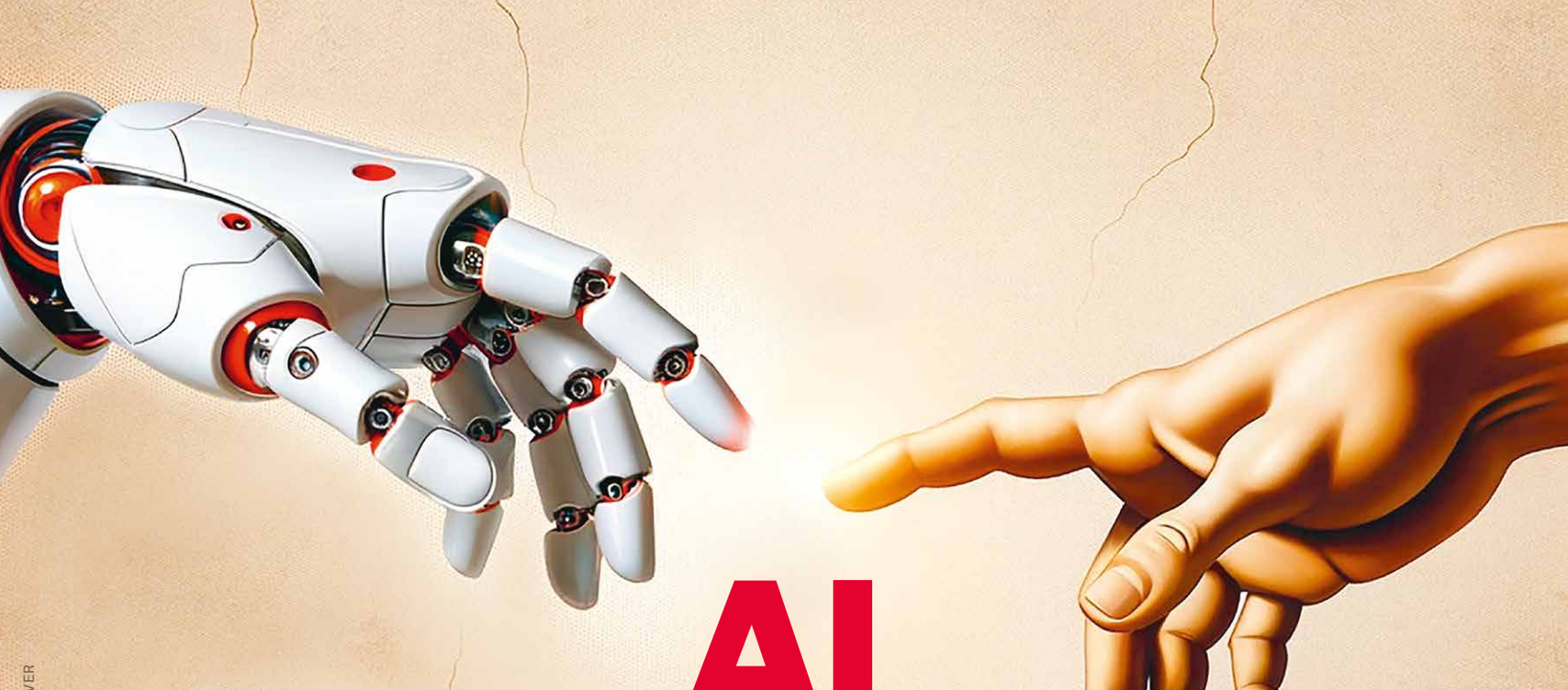

ILLUSTRATIE JANA HOLTHÖWER

AI moet een mens helpen, niet vervangen

CHARLOTTE VLEK

ONDERZOEK

WWW.RINEKEVERBRUGGE.NL

Van medische diagnoses tot autonome wapens in het Midden-Oosten: Artificial Intelligence (AI) neemt steeds meer beslissingen zelf, zonder dat er nog een mens aan te pas komt. Rineke Verbrugge, hoogleraar logica en cognitie aan de Faculteit Science and Engineering, vindt dat dat anders moet: AI zou een aanvulling moeten zijn op menselijke intelligentie, niet een vervanging ervan. Hoe? Dat onderzoekt ze samen met computerwetenschappers, psychologen en taalkundigen.

Hybrid Intelligence Centre

Computers kunnen razendsnel patronen herkennen, grote hoeveelheden data analyseren, en ze kunnen ons helpen objectief te blijven. Maar om de menselijke waarden en normen te behouden moet er altijd een mens *in de loop* zijn, vindt Verbrugge. 'Denk aan medische diagnoses: daar is AI heel goed in, bijvoorbeeld door beelden van MRI-scans te analyseren. Maar als het gaat om zaken van leven of dood zou een mens wel de uiteindelijke beslissing moeten nemen.'

De missie van het Hybrid Intelligence Centre – waarvan Verbrugge mede-oprichter is – is daarom AI-systemen te ontwerpen die kunnen fungeren als een extensie van ons menselijke verstand. Dat betekent dat het systeem niet zomaar een eindoordeel uitspuugt – een medische diagnose of de kortste route naar huis – maar kan samenwerken met en zich aanpassen aan mensen, dat verantwoorde keuzes kan maken en kan uitleggen waarom het die keuzes maakt. Dan zou een arts samen met het AI-algoritme een diagnose kunnen stellen, op basis van data én gesprekken met de patiënt.

'Als het gaat om leven of dood zou een mens wel de uiteindelijke beslissing moeten nemen'

Coöperatieve robotarm

'Het Hybrid Intelligence Centre doet al zijn onderzoek altijd met het oog op specifieke toepassingen', vertelt Verbrugge. 'We zijn nu vijf jaar bezig, over nog eens vijf jaar hopen we die toepassingen zo ver gebracht te hebben dat anderen het op de markt kunnen brengen.' De toepassingen lopen uiteen van een robotsuppoost die in een museum vragen van bezoekers kan beantwoorden, tot een robotarm die een chirurg bijstaat tijdens de operatie.

'Zo'n robotarm heeft een camera en maakt voor de chirurg close-up beeld dat op een beeldscherm getoond wordt. Dan is het handig als de robotarm anticipeert op wat de chirurg gaat doen, of even overlegt: "vind je het handig als ik straks even van de andere kant kijk?"' Ook zou zo'n robotarm kunnen inschatten wanneer een chirurg moe begint

te worden, en hem dan kunnen adviseren: 'misschien is het tijd je te laten vervangen'. 'Maar dat accepteert zo'n chirurg natuurlijk nooit!' lacht Verbrugge. En daarom is het belangrijk dat Hybrid Intelligence ook begrijpt hoe mensen in elkaar steken.'

Geleerde dwazen

'Computers zijn als geleerde dwazen: ze zijn ontzettend goed in slechts een smal domein, maar de vaardigheid om beweegredenen van mensen te doorgronden ontbreekt. Het is iets dat een kind van ongeveer vier jaar oud al wel kan: *theory of mind*, het kunnen redeneren over de gedachten, intenties en overtuigingen van een ander.' Verbrugge doet al bijna twintig jaar onderzoek naar het onderwerp.

Robot met theory of mind

Verbrugge bestudeert theory of mind vanuit een unieke combinatie van logica en cognitiewetenschap. Samen met computerwetenschappers, psychologen en taalkundigen achterhaalt ze hoe mensen theory of mind toepassen, en wanneer ze de mist in gaan. Het doel: door te begrijpen hoe mensen deze vaardigheid leren, zou je het ook aan een computer kunnen leren. Het er simpelweg in programmeren gaat namelijk niet, legt Verbrugge uit. 'Je zou een computer wel kunnen programmeren om over een mens te redeneren als die perfect logisch in elkaar steekt, maar we zijn als mensen nu eenmaal niet perfect. Verre van, zelfs.'

Mensen willen bijvoorbeeld niet altijd met elkaar samenwerken, of hebben niet altijd goede intenties. Dat is vooral ingewikkeld voor AI wanneer het in een gemengd team met mensen moet leren samenwerken. Bijvoorbeeld omdat de één graag promotie wil maken, waardoor een ander niet met diegene wil werken. 'Een AI die dat begrijpt, kan beter met die mensen samenwerken.'

Daarom onderzoeken promovendi van Verbrugge bijvoorbeeld hoe binnen een groep mensen zogenaamde 'common ground' kan ontstaan: de (on)geschreven regels waarvan iedereen op de hoogte is. Dat kan gaan over het gedrag van deelnemers aan een spel, fietsers bij een stoplicht dat alle kanten tegelijk op groen laat springen, of over privacy: wanneer is het sociaal acceptabel om een foto van een vriend(in) te delen op Facebook? En hoe kom je er eigenlijk achter wanneer een mens liegt, kun je dat ook aan een klein kind leren? En aan een computer?

Unieke studie Hybrid Intelligence

Verbrugge vindt het belangrijk ook buiten haar onderzoeksgroep jonge AI-onderzoekers te bereiken met de visie van het Hybrid Intelligence Centre. Als Director of Training and Eduction organiseert ze tweewekelijkse bijeenkomsten over Hybrid Intelligence voor promovendi van het onderzoekscentrum, en cursussen waar ook promovendi van daarbuiten welkom zijn. En de RUG krijgt binnenkort zelfs als eerste universiteit in de wereld een afstudeerrichting Hybrid Intelligence binnen de Masteropleiding AI. Verbrugge: 'Voor studenten die houden van programmeren, maar wel met oog voor de mens.'

FOTO LEONIE VON RIJTOK

Rineke Verbrugge (1965) is een pionier in het bouwen van bruggen tussen logica en cognitieve wetenschap. Ze behaalde in 1988 cum laude haar doctoraal en in 1993 haar PhD in Logica en Grondslagen van de Wiskunde aan de Universiteit van Amsterdam. Daarna was ze onderzoeker in Praag, Göteborg, aan het MIT in Cambridge (waar ze switchte naar kunstmatige intelligentie) en aan de VU in Amsterdam. Sinds 2002 is ze verbonden aan de RUG, waar zij in 2009 hoogleraar Logica en Cognitie aan het Bernoulli Instituut werd. Ze won vele awards en is lid van het KHMW en de KNAW. Ook is ze betrokken bij het Hybrid Intelligence Centre waarvan ze mede-oprichter is.

Sally-Anne taak

De klassieke test om theory of mind bij kinderen te onderzoeken bestaat uit een kort verhaaltje over Sally en Anne. Sally stopt een knikker in haar mandje, en gaat dan de kamer uit. Als ze weg is, haalt Anne de knikker uit het mandje en stopt die in een doosje. Dan komt Sally de kamer weer in. Een jong proefpersoontje krijgt nu de vraag: 'Waar denkt Sally dat haar knikker verstopt is?'

Een kind dat nog geen theory of mind heeft ontwikkeld, zal antwoorden dat Sally haar knikker in het doosje zal zoeken, want het kind weet zelf immers dat daar die knikker zit. Dat Sally dat niet weet, begrijpt ze nog niet. Een kind dat een zogenaamde eerste orde theory of mind heeft, zal zeggen dat Sally in haar mandje zal zoeken, daar heeft ze de knikker immers achtergelaten voor ze de kamer uit ging!

Tweede orde theory of mind: De opgave wordt een slagje moeilijker als de vraag niet gaat over wat Sally denkt, maar over wat Sally denkt dat Anne denkt. Bijvoorbeeld: waar denkt Sally dat Anne denkt dat de knikker is? Verbrugge legt uit dat theory of mind heel domein-specifiek is. Een verhaaltjestest lukt dan bijvoorbeeld wel in de tweede orde, maar bij een spelletje redeneren over wat de tegenspeler denkt dat jij denkt dat zij gaat doen, lukt nog niet.

Lees in dit fictieve verhaal wat er zou kunnen gebeuren als we AI een (te) grote rol geven in beslissingen die over ons leven gaan.

www.rug.nl/robot-doktersjas