Zoek een review

Mens en intelligente machine

PREMIUM

De waarde van samenwerking

Door **Dorine Schenk**
Freelanceredacteur
Feedback • 06-01-2021 06:00 •
V: 7.326, U: 1.027 26

Samen sterk

In 1997 versloeg IBM-supercomputer Deep Blue schaakgrootmeester Garri Kasparov, destijds wereldkampioen. Een computer had de mens verslagen. Acht jaar later organiseerde Playchess.com een online 'freestyle' schaaktoernooi. Alles was toegestaan; in teams spelen, computers laten schaken of werken in gemengde teams van mensen en computers. De uitkomst van het toernooi was verrassend. Niet de krachtigste computers of de beste grootmeesters wonnen, maar twee Amerikaanse amateurschakers die samenwerkten met computers.

Mens en machine bleken samen sterker dan een krachtige computer alleen en dit geldt niet alleen voor een potje schaak. Uit onderzoek bleek ook dat pathologen die samen met kunstmatig intelligente computersystemen borstkanker diagnosticeren, er minder vaak naast zitten dan een patholoog of AI alleen.

Frank van Harmelen

Dergelijke hybride samenwerkingen, tussen mens en machine, zijn volgens AI-onderzoekers Jurriaan van Diggelen van TNO en Frank van Harmelen van de VU de toekomst. Zij denken niet dat computers mensen compleet gaan vervangen. Van Harmelen: "Ik hoop dat we naar een wereld toegaan waarin we op allerlei vlakken op voet van gelijkheid zullen samenwerken met digitale partners."

Digitale partners zouden ons kunnen helpen om grootschalige problemen, zoals klimaatverandering, schaarste van goederen en pandemieën op te lossen door te voorkomen dat wij worden beperkt door de natuurlijke tekortkomingen van ons menselijke brein, zoals kortetermijndenken, moeite hebben met het bevatten van kansen en het denken in stereotypen.

Voordat we zover zijn, moeten mens en machine leren om elkaar te begrijpen en te vertrouwen. Nu is dat alleen [illegible] bleek uit het freestyleschaaktoernooi. Daaraan deden ook teams mee van

Inhoudsopgave

interfaces nodig hebben.

Jurriaan van Diggelen

Van Diggelen onderzoekt bij TNO hoe de samenwerking tussen mens en machine kan worden verbeterd. Hij werkt binnen het APPL.AI onderzoeksprogramma onder andere aan AI voor *urban search and rescue*. Hieronder vallen bijvoorbeeld robots die zoeken naar slachtoffers in ingestorte gebouwen die te gevaarlijk zijn voor mensen.

Daarnaast loopt in Nederland een project waarbij AI-onderzoekers samen met taalkundigen, psychologen en experts op het gebied van mens-machine-interacties werken aan AI-systemen die niet bedoeld zijn om mensen te vervangen, maar om met ze samen te werken. Dit gebeurt onder leiding van Van Harmelen aan het Hybrid Intelligence Centre. NWO kende vorig jaar een Zwaartekrachtsubsidie van bijna 20 miljoen toe aan deze samenwerking tussen zes Nederlandse universiteiten.

Met deze twee Nederlandse AI-onderzoekers bespreken we welke stappen nodig zijn voor een wereld waarin mensen op gelijke voet samenwerken met machines.

Robots die met hulp van TNO in een voor de mens gevaarlijke omgeving, een aardbevingsgebied, zijn ingezet. Mensen bleven weg van die plek, maar werkten op afstand wel samen, met machines.
eelding: Jurriaan van Diggelen (TNO) Coverafbeelding: Huang Yi Studio

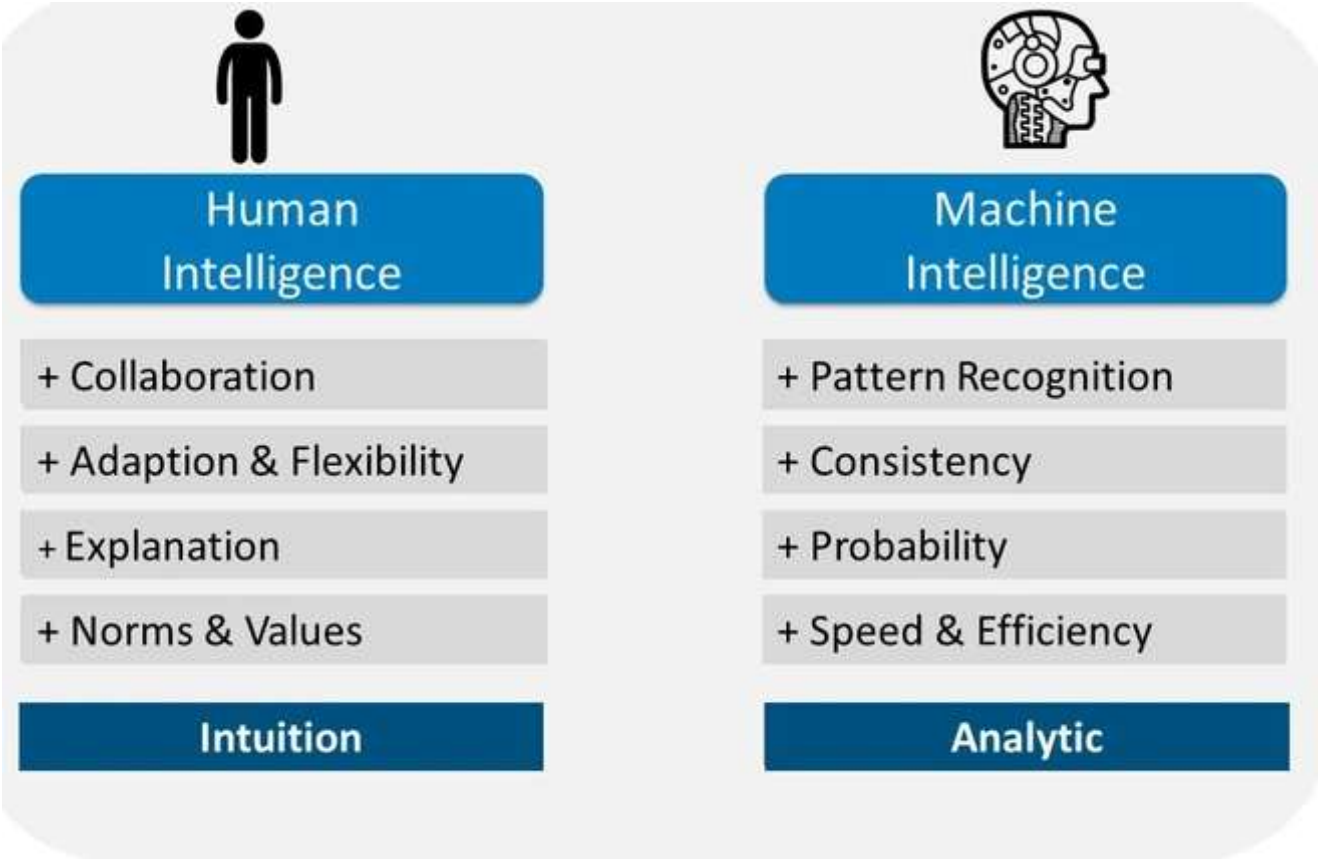
Inhoudsopgave

"Veel van het huidige AI-onderzoek gaat impliciet of expliciet over het vervangen van menselijke taken door intelligente machines", vertelt Van Harmelen. Denk bijvoorbeeld aan zelfrijdende auto's die taxichauffeurs overbodig maken en beeldanalysesoftware die ervoor zorgt dat radiologen gemist kunnen worden. Hierbij wordt ervan uitgegaan dat een computer intelligent is als hij een mens kan nadoen. Dat beeld heerst sinds de Turing-test uit 1950. Een machine slaagt voor deze test als hij in een gesprek een mens kan laten denken dat hij ook een mens is.

"Maar dit beeld klopt niet. Machine-intelligentie is een heel ander soort intelligentie dan die van mensen", vertelt Van Harmelen. Soms betekent dat dat machines ergens beter in zijn en soms lijken ze juist oliedom. Een bekend voorbeeld is IBM's supercomputer Watson, die in 2011 de Amerikaanse kennisquiz Jeopardy won. Watsons overwinning was een verrassing.

Gekke fouten

Nog verrassender waren de gekke fouten die de computer maakte. Zo kwam in de categorie 'Amerikaanse steden' de vraag: 'Van welke stad is de grootste luchthaven vernoemd naar een held uit de Tweede Wereldoorlog; en de op een na grootste naar een veldslag uit de Tweede Wereldoorlog?' Watson antwoordde: Toronto. Zelfs als je het antwoord niet wist, wist je waarschijnlijk dat het niet Toronto is. Watsons bouwers suggereerden naderhand dat de fout misschien ontstond doordat Watson niet begreep dat je bij het tweede deel van de vraag, na de puntkomma, uit de context moest afleiden dat 'de op een na grootste' naar een vliegveld verwees. En ter verdediging van Watson: er bestaan een paar (kleine) Amerikaanse steden die Toronto heten.



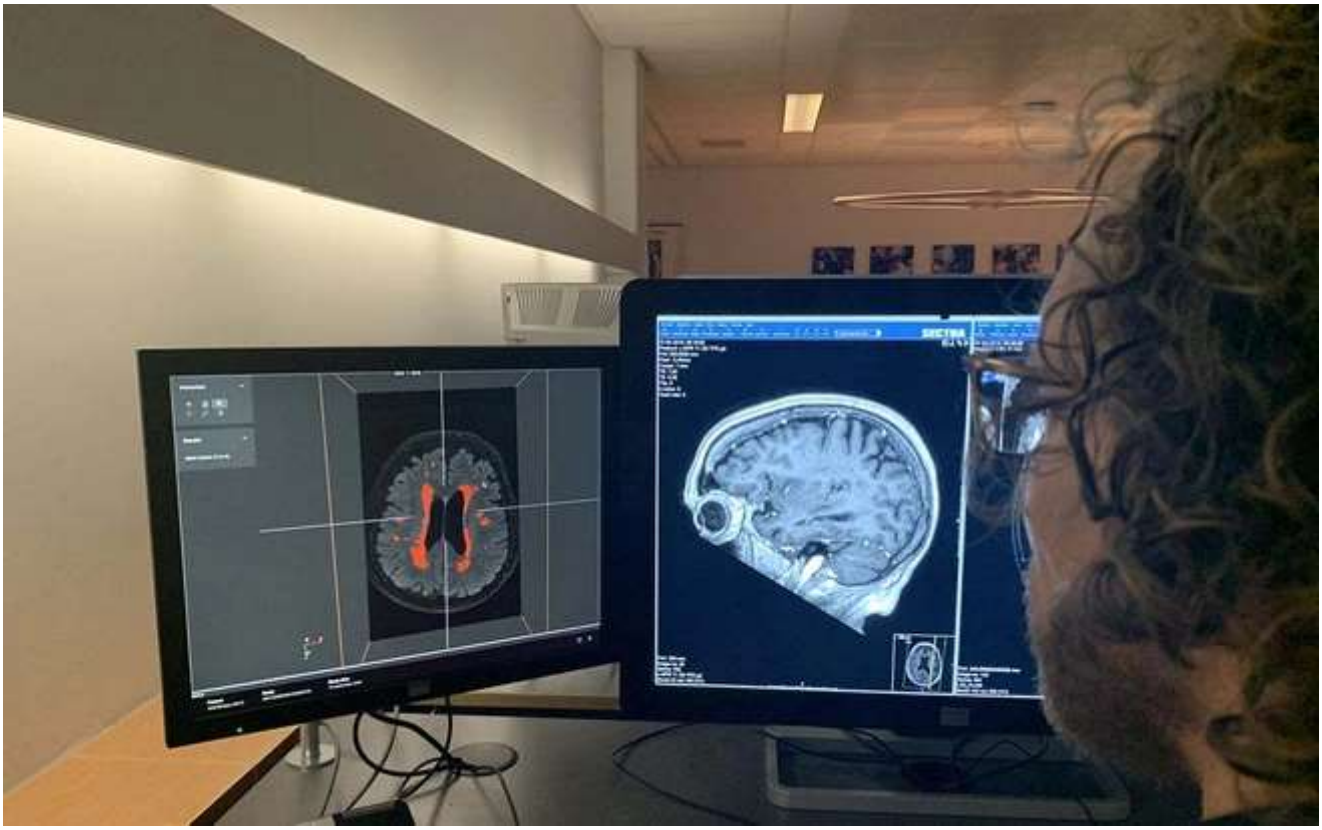
Mens en machine hebben dus andere vaardigheden. Computers kunnen razendsnel grote hoeveelheden data en optimalisatieproblemen oplossen. Ze zijn goed in het herkennen van

Inhoudsopgave

niet het antwoord is op een vraag over Amerikaanse steden met vliegvelden. Verder kunnen we samenwerken en passen we ons aan als omstandigheden veranderen. Van Diggelen: “Bovendien hebben mensen een moreel kompas. Dit is moeilijk te ontwikkelen in een computer die onbekend is met menselijke ervaringen als angst, schaamte en verdriet.”

Dit betekent dat soms computers en soms mensen bepaalde taken beter kunnen. “Het is bijvoorbeeld een heel slecht idee dat mensen nog steeds autorijden op de snelweg”, zegt Van Harmelen. “Onze reactietijd is honderden milliseconden. Dat is oerlangzaam. Bovendien worden we afgeleid door de radio of medepassagiers, of we vallen in slaap. Computers kunnen dat veel beter, waardoor auto’s veilig op tien centimeter afstand achter elkaar zouden kunnen rijden.” Dat is slecht nieuws voor bepaalde beroepsgroepen, zoals vrachtwagenchauffeurs.

Bij andere beroepen en taken zal juist samenwerking de beste oplossing zijn. Dit geldt bijvoorbeeld voor radiologen. Het herkennen van tumoren op röntgenfoto’s zullen AI’s op termijn waarschijnlijk beter kunnen, maar radiologen doen meer. Van Diggelen: “Ze zetten een therapie uit en hebben contact met de patiënt. Ik zie een slechtnieuwsgesprek door een computer niet gebeuren.” In een interview in NRC zei een radioloog hierover: “Radiologen verdwijnen niet, maar radiologen die geen kunstmatige intelligentie gebruiken, worden vervangen door hen die dat wel doen.”



UMC Utrecht ontwikkelde een platform voor AI-applicaties voor radiologen. Bron: UMC Utrecht

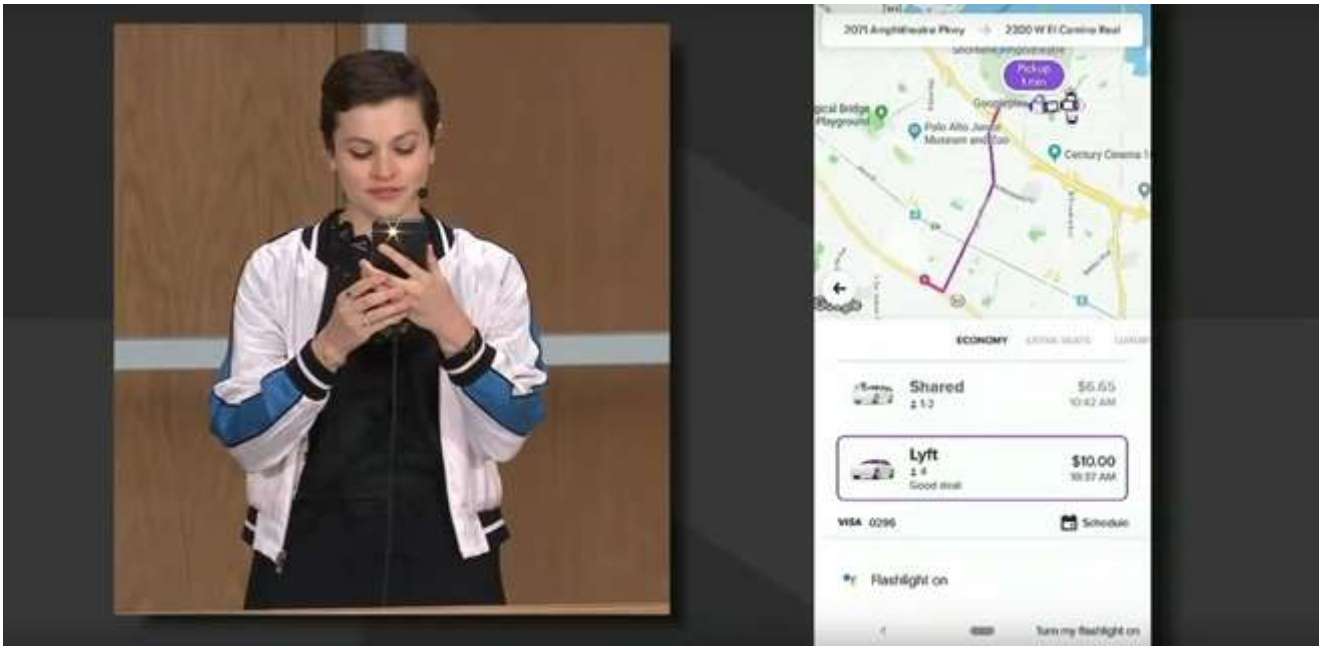
Vraag een computer: waarom?

Voor de samenwerking tussen mens en machine is wederzijds begrip nodig. Als een bewakingscamera iemand als [illegible]grijk om te weten hoe en waarom een AI tot die conclusie komt. Hoe zorgen we dat [illegible]n? Daarvoor zullen computers moeten beschikken over een aantal vaardigheden

Inhoudsopgave

Mensen gebruiken, vaak onbewust, sociale vaardigheden om elkaars doelen en kennis te peilen. Zo wist Frank van Harmelen wat voor stuk ik wilde schrijven en welke achtergrondkennis ik heb, waardoor hij kon schatten welke delen van de uitleg over zijn onderzoek hij kon overslaan. Dit vermogen om het perspectief van anderen te begrijpen, wordt *theory of mind* genoemd. Kort gezegd: ik weet wat jij weet en jij weet dat ik weet wat jij weet. Onderzoekers van het Nederlandse Hybrid Intelligence Centre werken aan een manier om computers te voorzien van een theory of mind. De bedoeling is dat een AI tijdens een dialoog met een mens het model van de gesprekspartner automatisch aanpast. Zoals wij ook onze theory of mind over een ander aanpassen op basis van antwoorden of bijvoorbeeld bepaald jargon dat hij of zij gebruikt.

Daarnaast wordt gewerkt aan manieren om de communicatiemogelijkheden uit te breiden. Een gesprek met bestaande virtuele assistenten, zoals Siri en Alexa, is beperkt tot een enkele iteratie: je stelt een vraag en krijgt een antwoord. Reageren op een vervolgvraag of een gesprek voeren waarbij voor- en tegenargumenten worden uitgewisseld, kunnen ze niet. Nog niet, want er bestaan formele theorieën over argumentatiestructuren die te vatten zijn in formele logica, vertelt Van Harmelen. “Die theorieën zijn goed genoeg gedefinieerd om ze te beschrijven in computercode, zodat een AI ermee overweg kan. Dit moet het binnenkort mogelijk maken om coherente gesprekken te hebben met een computer, in plaats van een beperkt vraag-en-antwoordspelletje.”



Google demonstreerde tijdens Google I/O 2019 Continued Conversation voor zijn Assistent, maar daarbij gaat het nog steeds om een reeks opdrachten, niet om een coherente conversatie.

Naast een goed gesprek voeren met je digitale collega wil je ook begrijpen waarom hij tot bepaalde conclusies komt. “Een samenwerkende machine moet haar resultaten of antwoorden kunnen uitleggen”, zegt Van Diggelen. “Veel beeldherkenningsalgoritmen kunnen dat niet, waardoor het lastig is de betrouwbaarheid ervan op waarde te schatten.”

Dit vakgebied, genaamd explainable AI, is bijvoorbeeld relevant voor kunstmatige neurale netwerken, geïnspireerd op de werking van de hersenen. Deze netwerken, die bestaan uit kunstmatige neuronen die onderling verbonden

een vraag stelt, zoals: wat zie ik op deze foto? Als je dan echter vraagt waarom,

Inhoudsopgave

antwoord geeft op de vraag: waarom?”

Black box

Het is ook geen eenvoudige uitdaging. Neurale netwerken zijn een black box, we weten niet precies hoe ze tot een conclusie komen. Ze bestaan uit enkele tot tientallen lagen met een paar duizend tot enkele miljoenen of zelfs een paar miljard kunstmatige neuronen en verbindingen. Al deze verbindingen en neuronen zijn parameters die tijdens het trainen geoptimaliseerd zijn om het goede antwoord te geven. In die brei van getweakte parameters kan niemand terugvinden waarom je een bepaald antwoord kreeg. “Dit is optimalisatie in een miljarddimensionale ruimte”, zegt Van Harmelen. Informatie over de neuronen die in de hersenpan van je ex vuurden, is ook nutteloos als je wilt begrijpen waarom hij of zij het uitmaakte.

In het geval van een getraind AI-systeem zou het nuttiger zijn als je bijvoorbeeld een antwoord terug kunt leiden naar de concrete voorbeelden waarop de AI getraind is. Dat is nu nog niet mogelijk, maar volgens Van Harmelen wordt er gewerkt aan verschillende technieken om dit voor elkaar te krijgen.

Er zijn ook andere manieren om een kunstmatig neuraal netwerk een verklaring te ontfutselen. Je kunt bijvoorbeeld een AI zijn keuzes achteraf laten beredeneren. Van Harmelen: “Je laat de computer dan, onafhankelijk van hoe hij tot het antwoord gekomen is, argumenten verzinnen waarom het een redelijk antwoord zou zijn geweest, door in eigen of andere bronnen te zoeken.” Dat lijkt misleidend, maar uit psychologisch onderzoek blijkt dat dit de manier is waarop mensen keuzes verklaren. Vaak is het besluit om bijvoorbeeld een bepaalde trui te kopen tot stand gekomen door onbewuste beïnvloeding, bijvoorbeeld door reclame. De meeste mensen zullen echter een min of meer verzonnen, achteraf beredeneerde verklaring geven. Bijvoorbeeld dat het model goed bij de rest van hun garderobe past. Van Harmelen: “We kunnen computers op een vergelijkbare manier achteraf een overtuigend verhaal laten construeren, al was dat niet de causale manier waarop zij aan de manier kwamen.” Voor veel samenwerkingen voldoet een dergelijke verklaring.

Ten slotte zijn er onderzoekers die proberen van de black box van AI’s een glass box te maken door pogingen te doen om te traceren hoe dat neurale netwerk tot een antwoord kwam. Mensen proberen dan bijvoorbeeld alleen naar de laatste paar lagen van het neurale netwerk te kijken. Ze hopen dat de laatste stappen die de AI genomen heeft, het antwoord op een voor ons begrijpelijke manier verklaren. De kunst is daarbij om enerzijds het netwerk zo eenvoudig te maken dat je het kunt uitleggen en het tegelijk ingewikkeld genoeg te maken om het goed te laten presteren.

De hierboven beschreven technieken gingen over vaardigheden die een AI moet verwerven, maar mensen kunnen ook wat doen om beter samen te werken met computers. “Mensen maken machines antropomorf en denken dat ze net zo redeneren als mensen”, zegt Van Diggelen. “Dat gaat mis. Als er meer hybride samenwerkingen ontstaan, moeten we misschien meer leren over wat ik maar even AI-psychologie noem, zodat we beter leren schatten wat AI-systemen wel en niet kunnen. Het is bijvoorbeeld niet onbelangrijk om te weten wat je zelfrijdende auto doet als er plots een boom op de weg ligt en de stoep nog wel vrij is. Machines zullen zich moeten verdiepen in de mens, maar nind vormen over de machine, zodat je snapt wat machines doen en kunt

Inhoudsopgave

Hoe zien de AI-onderzoekers de toekomstige samenwerking tussen mens en machine voor zich? Van Diggelen: “Het zou eruit kunnen zien als een bedrijfssituatie waarin de mens de rol aanneemt van manager en de AI die van medewerker. Een manager bepaalt in grote lijnen wat er gebeurt en stuurt medewerkers aan, maar laat ze ook vrij in de manier waarop ze hun projecten uitvoeren. De medewerker klopt niet voor elk wissewasje aan bij de manager, maar meldt het wel als er echt een probleem is. Naar die situatie willen toe we met AI-systemen. Managers houden ook regelmatig een functioneringsgesprek. Iets vergelijkbaars is nodig, zodat we grip houden op wat AI-systemen wel en niet kunnen.”

Het doel van het Hybrid Intellligence Centre van Van Harmelen is ambitieus. In het *moonshot*-project van het Hybrid Intellligence Centre komen alle technieken die nodig zijn voor hybride intelligentie, waarbij mens en AI op gelijke voet samenwerken, samen in een digitale wetenschappelijke collega. Van Harmelen: “We willen een AI-programma ontwikkelen dat met ons of met andere wetenschappers samenwerkt in alle stappen van een wetenschappelijk onderzoek: van literatuur lezen, een hypothese opstellen en een experiment ontwerpen tot het analyseren van de data en het schrijven van een artikel waarin de resultaten gerapporteerd worden.”

Een digitale coauteur zou bijvoorbeeld kunnen helpen met het lezen van de literatuur. Van Harmelen: “Ik wil op maandagochtend mijn kantoor binnenkomen en van de computer een relevante samenvatting krijgen van de duizenden artikelen die hij dat weekend gelezen heeft.”

Bij symbolische AI leert een kunstmatige intelligentie grofweg op basis van door mensen leesbare informatie.
Bij subsymbolisch is dat door onbewerkte informatie in te voeren voor analyse door de AI.

Een aantal bouwstenen van zo'n kunstmatige coauteur is er al. Er bestaan computerprogramma's die [illegible]uur. En voor het analyseren van experimentele data worden al regelmatig AI's

Inhoudsopgave

in die data gebeurt. Van Harmelen: “AI’s die hypotheses kunnen ontwikkelen, zijn minder ver ontwikkeld. Daar werken we nu aan, in samenwerkingsprojecten met medische en historische wetenschappen.” De lastigste bouwsteen is een computerprogramma ontwikkelen dat een experiment kan ontwerpen.

“Nu zijn deze bouwstenen nog autistische programma’s die geen idee hebben van de context waarin ze functioneren”, zegt Van Harmelen. Die losse onderdelen aan elkaar koppelen, zodat ze onderling en met menselijke collega’s kunnen communiceren, wordt de grootste stap. “We denken echter dat het in de komende acht jaar haalbaar is om een wetenschappelijk artikel te schrijven waarbij we met droge ogen een AI als coauteur kunnen opvoeren.”

En daarna? “Als dit zich verder ontwikkelt, denk ik dat de samenwerking tussen mens en machine op onverwachte plekken zal ontstaan; zowel in werkverband als in ons persoonlijke leven”, besluit Van Harmelen. “Mijn hoop is dat de samenwerking met computers kan helpen om te compenseren voor onze menselijke tekortkomingen. Door bijvoorbeeld te voorkomen dat sollicitatiecommissies onbewust kiezen voor iemand die op henzelf lijkt of door ons te helpen om gezondere keuzes in ons leven te maken.”

De VU Amsterdam zet onder andere de Pepper-robot 'Leolani' in, bij onderzoek naar taal en communicatie door computers en robots.

Multipage layout

Inhoudsopgave