MEERWAARDE VAN MENS MET MACHINE

Door Bennie Mols Beeld Shutterstock en Ivar Pel

Kunstmatige intelligentie ontwikkelen die goed met mensen kan samenwerken. Dat is het doel van het NWO Zwaartekrachtprogramma 'Hybrid Intelligence: Augmenting Human Intellect'.

Computers zijn veel beter dan mensen in strategische spellen als schaken en Go, maar mensen hebben veel meer gezond verstand dan computers. Robots kunnen perfect de hele dag grote auto-onderdelen in elkaar zetten, maar ondanks de jarenzestigfantasieën over Rosie de Robot zijn mensen nog steeds veel handiger dan robots in het doen van het huishouden. Voor veel toepassingen in de echte wereld is de combinatie van mens en machine krachtiger dan een van beide afzonderlijk. Dat is de achterliggende filosofie van het onderzoeksprogramma Hybrid Intelligence dat onlangs een NWO Zwaartekrachtsubsidie kreeg van twintig miljoen euro voor tien jaar. Bij het programma zijn in totaal zo'n dertig onderzoekers van zes Nederlandse universiteiten betrokken.

Medebedenker en programmaleider is Frank van Harmelen, hoogleraar Kunstmatige intelligentie aan de Vrije Universiteit Amsterdam. 'Veel van de kunstmatige intelligentie van dit moment is gericht op machines die autonoom functioneren en mensen vervangen. Ons doel is juist om kunstmatige intelligentie technisch geschikt te maken om met mensen samen te werken. Om dat doel te halen werkt een groot deel van de onderzoekers binnen kunstmatige intelligentie in Nederland met elkaar samen. Dat vind ik heel bijzonder aan ons programma.'

De onderzoekers schetsen drie toekomstscenario's voor de zorg, het onderwijs en de wetenschap om hun doelen concreet te maken. Van Harmelen: 'In de zorg is ons doel om robots samen met verpleegkundigen te laten zorgen voor kinderen met kanker. In het onderwijs willen we kijken hoe robots en menselijke docenten samen beter onderwijs kunnen geven. En in de wetenschap hebben we echt een moonshot geformuleerd: het publiceren van een wetenschappelijk artikel waarvan de computer coauteur is. Dat is het moeilijkst te halen doel.'

Vier lijnen

Om deze doelen te halen, werken de onderzoekers in vier onderzoekslijnen, samengevat in de afkorting CARE: Collaborative, Adaptive, Responsible en Explainable. In de onderzoekslijn Collaborative draait het om de samenwerking tussen mens en machine. Van Harmelen: 'In de kunstmatige intelligentie van dit moment onderzoekt men vooral de samenwerking tussen machines onderling: multi-agent-systemen heet dat vakgebied. De samenwerking tussen mens en machine stelt ons echter voor heel nieuwe uitdagingen. Zo zouden machines dan ook moeten weten wat mensen wel en niet weten. Machines moeten een *theory of mind* hebben, zoals dat heet. Dat hebben ze nu nauwelijks.'

Daarnaast draait in de huidige kunstmatige intelligentie veel om het machinaal leren van grote hoeveelheden data. Maar de wereld verandert voortdurend en na iedere verandering heeft een getrainde machine vaak geen antwoord meer op de nieuwe uitdagingen. 'Voor het omgaan met zo'n dynamische wereld moeten machines dus ook adaptief zijn', zegt Van Harmelen. Dat is de tweede onderzoekslijn. 'Kunstmatige intelligentie heeft daarvoor niet alleen data nodig, maar ook kennis van de wereld. Beide aanpakken willen we combineren.'

De derde onderzoekslijn, Responsible, draait om het bijbrengen van normen en waarden aan machines. Van Harmelen: 'We willen wel dat onze machines zich gedragen naar de normen en waarden die wij als mensen belangrijk vinden. Een machine moet bijvoorbeeld leren dat hij niet mag liegen. En een robot die voor de klas staat zou ook moeten weten hoe hij zich tegenover zijn leerlingen moet gedragen. Net als mensen moeten machines leren socialiseren.'

De vierde en laatste onderzoekslijn, Explainable, heeft als doel om machines te laten uitleggen waarom ze een bepaalde beslissing willen nemen of hebben genomen. Veel lerende systemen van tegenwoordig werken als een zwarte doos. Maar om met mensen samen te werken zullen machines ook moeten kunnen uitleggen wat ze willen en waarom.

Volgens Van Harmelen loopt Nederland voorop in het onderzoek naar hybride intelligentie: 'Je ziet op sommige plekken in de wereld soortgelijke initiatieven ontstaan, bijvoorbeeld aan Stanford in de VS, maar wij waren er vroeg bij. Ook ligt ons onderzoeksplan sterk in lijn met de filosofie die de EU heeft over kunstmatige intelligentie.'

Van Harmelen:

'Ons doel is om kunstmatige intelligentie technisch geschikt te maken om met mensen samen te werken'

Indenken in de ander

Hoogleraar Logica en cognitie Rineke Verbrugge van de Rijksuniversiteit Groningen wordt binnen het programma verantwoordelijk voor het onderzoek naar hoe machines een theory of mind kunnen krijgen. Verbrugge: 'Ik definieer theory of mind als je kunnen indenken wat een ander denkt, gelooft, weet en bedoelt. In de sociale interactie gebruiken mensen dat voortdurend. Ik weet dat jij weet dat Amsterdam de hoofdstad is van Nederland. Maar ik weet ook dat jij weet dat ik weet dat Amsterdam de hoofdstad is van Nederland.'

Mensen ontwikkelen die vaardigheid geleidelijk. Ergens tussen het derde en vierde levensjaar leren kinderen begrijpen wat iemand anders wel of niet weet. Rond hun zesde of zevende leren kinderen redeneren over wat ze denken dat anderen over hen denken. En zo rond de tienjarige leeftijd begrijpen kinderen wat het betekent als iemand anders iets expliciet belooft.

Het onderzoek van Verbrugge gaat een belangrijke rol spelen in de onderzoekslijn Collaborative. Verbrugge: 'Bij samenwerken hoort een collectieve intentie. Je wilt samen een doel bereiken. Dan heb ik bijvoorbeeld de intentie dat jij meewerkt, maar ik wil ook dat jij weet dat ik wil meewerken. Als computers in het team zitten dan zullen ze ook moeten begrijpen hoe de menselijke theory of mind werkt. Wat weet een ander teamlid en wat niet? Wat zijn de wensen van andere teamleden? Momenteel kunnen computers dat alleen in geïdealiseerde spelomgevingen met duidelijke spelregels en duidelijke ja of nee beslissingen. Maar in de echte wereld schieten computers nog tekort. Neem het conflict tussen Israël en de Palestijnse Autoriteit. Daarin spelen emoties een grote rol, net als de gecompliceerde geschiedenis. Huidige computers kunnen daar niet mee omgaan. Hetzelfde geldt voor het begrijpen van lichaamstaal. Mensen kijken naar een ander om er achter te komen wat iemand denkt. Computers zijn daar slecht in.'

Computer als coauteur

Stel dat een computer over tien jaar inderdaad coauteur van een wetenschappelijk artikel wordt, wat moet de machine dan beheersen aan theory of mind? Verbrugge: 'Dan moet de computer vragen kunnen beantwoorden over het artikel. Hij moet ook kritiek kunnen pareren. Of kunnen zeggen: "Sorry, dat heb ik verkeerd geïnterpreteerd, een andere interpretatie ligt toch meer voor de hand." De computer mag alleen maar als coauteur bij het artikel als hij volledig dezelfde verantwoordelijkheden heeft als de menselijke coauteurs.' Daarnaast zou de computer ook in staat moeten zijn tot meta-redeneren: hij moet inzicht hebben in waarin hij zelf wel en niet goed is en waarin zijn teamleden wel of niet goed zijn. Verbrugge: 'Het zou mooi zijn als de computer zelf bedenkt dat die en die wetenschappers elkaar nog niet kennen maar gezien hun expertises wel eens heel goed zouden kunnen samenwerken aan het artikel.'

Naast haar wetenschappelijk-inhoudelijke taak wordt Verbrugge ook voor een dag per week verantwoordelijk voor de educatie en training binnen het programma. 'We willen de volgende generatie voldoende bagage meegeven, zowel de studenten en promovendi, als de postdocs en universitair docenten. En die bagage omvat meer dan alleen wetenschap. Promovendi die later het bedrijfsleven willen ingaan willen we de juiste training geven, maar ook promovendi die hoogleraar willen worden. En omdat kunstmatige intelligentie een steeds grotere impact op de samenleving gaat krijgen, willen we onze jonge onderzoekers ook voorbereiden op het communiceren van hun wetenschap naar een algemeen publiek. En natuurlijk moeten ze leren nadenken over de ethische en sociaal-maatschappelijke gevolgen van hun werk.'

Verbrugge:

'Mensen kijken naar een ander om er achter te komen wat iemand denkt. Computers zijn daar slecht in'

Meer informatie: www.hybrid-intelligence-centre.nl

ABSTRACT

The research program Hybrid Intelligence has received an NWO Gravitation grant of twenty million euro for ten years. Its main aim is to develop AI systems that can cooperate with people. To make this aim concrete the program sketches three future scenarios. In healthcare the aim is to develop a team of care robots and human caretakers to help children with cancer in a specialised hospital. In education the aim is to develop teams of robots and teachers to improve education. Finally, in science the aim is to publish a scientific article with a computer as a co-author.

PROGRAMMA IN VOGELVLUCHT:

Budget: Twintig miljoen euro.
Looptijd: Tien jaar.

Betrokken universiteiten: Vrije Universiteit Amsterdam (penvoerder), Universiteit van Amsterdam, Technische Universiteit Delft, Universiteit Leiden, Universiteit Utrecht, Rijksuniversiteit Groningen. Onderzoekslijnen: Vier aspecten van kunstmatige intelligentie, samengevat in de afkorting CARE: Collaborative, Adaptive, Responsible en Explainable. Praktijkdoelen: In de zorg – robots die samen met verpleegkundigen samenwerken in ziekenhuizen voor kinderen met kanker. In het onderwijs – teams van robots en docenten om samen beter les te geven. In de wetenschap – een gepubliceerd wetenschappelijk artikel waarvan een computer coauteur is.