

Computer legt het af tegen peuterintuïtie

Hoe kan het toch dat computers op bovenmenselijk niveau schaken en go spelen, maar qua gezond verstand nog niet in de buurt komen van driejarigen?

Tekst **Bennie Mols** Foto **Annabel Oosteweeghel**



Een man met een stapel papier in zijn handen loopt naar een gesloten kast toe. In de hoek van de kamer kijkt een klein jongetje toe. De man botst tegen de kast op, doet een paar stappen terug, loopt weer naar voren en botst voor de tweede keer tegen de kast op. Weer doet hij een paar stappen terug. Dan loopt het jongetje naar de kast, opent de deuren en kijkt omhoog naar de man. De man loopt weer naar de kast en terwijl het jongetje omhoog kijkt naar een boekenplank, legt de man de stapel papier in de kast.

De video van deze scène is afkomstig van de psychologen Felix Warneken en Michael Tomasello. Zij deden onderzoek naar altruïsme bij jonge kinderen.

De Amerikaanse cognitiewetenschapper en onderzoeker van kunstmatige intelligentie Josh Tenenbaum, verbonden aan het MIT in Massachusetts, gebruikt dezelfde video juist om de beperkingen van de huidige kunstmatige intelligentie te illustreren. Tenenbaum zei afgelopen juli op een congres over kunstmatige intelligentie: „Het jongetje is anderhalf jaar oud en heeft deze situatie nog nooit eerder gezien. Hij begrijpt de bedoeling van de man en hij begrijpt de functie van de kast. Denk je eens in wat het jongetje allemaal moet weten om te doen wat hij doet. Geen enkele computer of robot kan wat dit jongetje kan.”

Computers zijn in de afgelopen jaren heel goed geworden in het herkennen van patronen, bijvoorbeeld om voorwerpen of gezichten in foto's te herkennen. Recent zijn er diverse doorbraken geweest: zo slaagde het computerprogramma AlphaZero er in januari voor het eerst in om zowel schaken, go als shogi (Japans schaken) op bovenmenselijk niveau te leren spelen, zonder enige andere voorkennis dan de basale spelregels. Maar menselijke intelligentie is meer dan patroonherkenning. Waarom is het voor computers nog moeilijk om het niveau van een peuter te halen?

Tenenbaum: „Intelligentie gaat ook over het modelleren van de wereld. Daar hoort bij: verklaren en begrijpen wat we zien, verbeelden wat we zouden kunnen zien maar nog niet hebben gezien, problemen oplossen, acties plannen en nieuwe modellen ontwikkelen terwijl we meer over de wereld leren. We zijn nog ver weg van kunstmatige intelligentie die de wereld zo flexibel en diep kan modelleren als mensen dat kunnen. Maar we hebben tenminste één weg om daar te komen, en dat is *reverse engineering* hoe die capaciteiten in het menselijk brein werken.”

Tenenbaum zelf ontwikkelt computermodellen voor wat hij intuïtieve natuurkunde en intuïtieve psychologie noemt. In de eerste twaalf maanden leren baby's bijvoorbeeld dat voorwerpen niet zomaar ophouden te bestaan als je ze niet meer ziet, dat ze niet zomaar naar een andere plek teleporteren en dat ze niet door obstakels heen kunnen. Ze leren de driedimensionale ruimte te begrijpen. Ze leren de concepten van zwaartekracht, traagheid en van oorzaak en gevolg. Dat hoort allemaal bij intuïtieve natuurkunde.

Bij intuïtieve psychologie hoort het leren dat mensen hun doelen meestal met zo min mogelijk inspanning proberen te bereiken. Dat weten drie maanden oude baby's al. En tien maanden oude baby's kunnen al een soort kosten-baten-analyse maken over hoe graag mensen een bepaald doel proberen te bereiken.

Reflexmachines

Om nog beter te begrijpen waarom het voor kunstmatige intelligentie nog steeds zo moeilijk is om te doen wat het anderhalf jaar oude jongetje doet, ga ik met de video op mijn laptop langs in Nijmegen, bij twee Nederlandse experts: cognitie-

wetenschapper Iris van Rooij en hoogleraar kunstmatige intelligentie Marcel van Gerven, beiden verbonden aan de Radboud Universiteit in Nijmegen. Allebei kennen ze het werk van Tenenbaum goed, maar kijken er vanuit een ander perspectief tegenaan: Van Rooij als cognitiewetenschapper, en Van Gerven als neurowetenschapper en onderzoeker en ontwikkelaar van kunstmatige intelligentie.

Van Gerven hield in september zijn oratie onder de titel ‘Menselijke machines’. Nadat hij het filmpje heeft gezien, zegt hij: „Het jongetje demonstreert empathie. Het kan zich verplaatsen in de man. Het kind heeft als het ware een intern model van andere mensen, van zichzelf en ook van de fysieke wereld om hem heen.”

Die interne modellen ontbreken in intelligente computers en robots. „In de kunstmatige intelligentie hebben we momenteel hele geavanceerde reflexmachines, die snel een voorwerp op een foto kunnen herkennen of snel een stuk tekst vertalen, maar zonder dat ze echt weten waar de tekst over gaat of echt weten wat er op de foto is te zien. Ze missen intentionaliteit.”

Van Rooij valt hetzelfde op, maar ze verwoordt het anders: „Het kind leidt uit het niet lukken van een handeling van de man af wat zijn bedoeling is. Dat doel zelf is echter een onobserveerbare mentale toestand. Dat maakt het zo fascinerend. De man kan een heleboel verschillende doelen hebben. Misschien wil hij de stapel op de kast leggen. Het jongetje zou de stapel ook zelf in de kast kunnen leggen, maar dat doet hij niet. Hij haalt alleen een obstakel voor de man weg.”

Van Rooij noemt het afleiden van het juiste doel uit wat het kind ziet „een probleem van computationele complexiteit”. „Daarmee bedoel ik dat als je dit probleem uitleen gaat rafelen in instructies voor een machine, je een rekenkundig probleem krijgt met zoveel mogelijkheden dat het zelfs met alle computers in de wereld nog niet exact is op te lossen. Veel van onze cognitieve vaardigheden, zoals waarnemen, plannen, taal leren, beslissen, categoriseren en analogieën leggen, lijken in de context van de echte wereld diezelfde rekenkundige complexiteit te hebben. Ons brein kan het, hoewel we nog niet precies weten hoe, maar onze machines lopen vaak aan tegen die rekenkundige complexiteit.”

Gezond verstand

Aan de andere kant zou je kunnen zeggen dat een spel als go toch ook rekenkundig heel complex is, maar dat computers daarin inmiddels veel beter zijn geworden dan mensen. Hoe zit dat dan?

„Dat is een grootse prestatie van de computer”, reageert van Rooij, „maar er zijn grote verschillen tussen de wereld van het go-spel en de echte wereld. In go zijn de regels precies bekend, is het doel

van spel precies bekend, zien beide spelers alle informatie en is er geen enkele onvoorspelbaarheid. De echte wereld is onvoorspelbaar en vaak is helemaal niet duidelijk welk probleem iemand moet oplossen. Ook is het aantal mogelijke situaties in principe oneindig, ook al maakt ieder mens er in zijn leven maar een eindig aantal mee.”

Dan is er nog een tweede probleem, zegt Van Rooij, dat het jongetje wel weet op te lossen, maar computers en robots nog niet. Dit is het zogeheten frame-probleem, dat diverse filosofen van kunstmatige intelligentie, onder wie Jerry Fodor en Daniel Dennett, al decennia geleden hebben beschreven. De kern ervan is de vraag hoe je bepaalt welke informatie in een gegeven situatie relevant is en welke informatie niet relevant is. Dat is een nog onopgelost fundamenteel probleem. Voor het jongetje is de kleur van de kast niet relevant, maar wel het feit dat de kast twee dichte deuren heeft die je kunt openen.

„Ons gezond verstand stelt ons in staat om adequaat te reageren op situaties die we nooit eerder hebben gezien”, zegt van Rooij. „Je zou de situatie visueel heel anders kunnen maken, terwijl het kind toch ziet dat de man hetzelfde doel heeft. Het gaat dus niet om eigenschappen die je aan de buitenkant kunt zien, maar om abstracte concepten. Het kind weet bijvoorbeeld ook dat de kast hol is en niet masief, en dat er waarschijnlijk planken in zitten waar je iets op kunt leggen.”

Optische illusies

Het feit dat het jongetje van anderhalf jaar iets kan wat computers en robots niet kunnen, en dat computers beter schaken en go spelen dan mensen, illustreert dat mensen en machines heel andere problemen oplossen. Van Rooij: „Dat zie je bijvoorbeeld op het gebied van visuele waarneming. Machines en mensen maken hier heel andere fouten. Een computer die normaliter perfect een banaan kan herkennen, denkt ineens dat die banaan een broodrooster is wanneer je een rare sticker naast de banaan legt. Geen kind maakt die fout, terwijl ons brein wel ten prooi valt aan allerlei andere optische illusies die computers niet herkennen.”

De intelligentie van huidige lerende machines is gebaseerd op neurale netwerken die zeer losjes geïnspireerd zijn op het menselijk brein. Zo'n neuraal netwerk bestaat uit kunstmatige neuronen die verdeeld worden in lagen. De neuronen worden met elkaar verbonden en de sterkte van die verbindingen verandert tijdens het leerproces. Alhoewel neurale netwerken zeer goed zijn in het oplossen van specifieke taken zijn ze nog ver verwijderd van de flexibele en adaptieve manier waarop ons eigen brein willekeurige problemen kan oplossen. Dat is het verschil tussen kunstmatige intelligentie in enge zin (*narrow AI*) en mensachtige

kunstmatige intelligentie, die veel algemener is (*general AI*).

Van Gerven: „In theorie is zo'n netwerk heel krachtig, maar aan de andere kant staan onze huidige neurale netwerken nog heel ver van ons eigen brein vandaan. Het brein heeft op een evolutionaire tijdschaal gespecialiseerde substructuren ontwikkeld die ons tezamen in staat stellen om te overleven in complexe en onzekere situaties. Daarnaast heeft ons brein ook allerlei gespecialiseerde neuronen. Zulke mate van detail hebben de huidige neurale netwerken nog lang niet. Het overbruggen van het gat tussen ons eigen brein en de neurale netwerken die in de AI gebruikt worden kan de sleutel zijn tot slimmere machines.”

Van Rooij vindt dat de recente successen van lerende computers de kunstmatige intelligentie als vakgebied hebben afgeleid van fundamenteelere problemen: „De huidige successen zijn behaald in domeinen die veel beperkter zijn dan de echte wereld. Maar het gaat juist mis bij het opschalen naar de echte wereld. Als cognitiewetenschapper zeg ik dan: de kunstmatige intelligentie begint nu op fundamentele problemen te stuiten die we al lang kennen uit het werk van filosofen en cognitiewetenschappers, zoals het frame-probleem of het probleem om causale verbanden te leggen tussen gebeurtenissen.”

Huidige lerende computers kunnen weliswaar correlaties in data ontdekken, maar niet automatisch de causale verbanden. Uit het aantal verdrinkingen per seizoen en de ijsjesverkoop per seizoen zou een computer concluderen dat er een sterk verband is tussen het aantal verdrinkingen en het aantal verkochte ijsjes. Dat die twee geen oorzakelijk verband met elkaar hebben, maar veroorzaakt worden door de hoogte van de buitentemperatuur, weet de computer niet.

Megawatts aan energie

Ook op hardware-niveau bestaan er nog grote verschillen tussen mens en machine. Het menselijk brein verbruikt slechts twintig watt aan energie. Supercomputers als Watson (die in 2011 de tv-quiz Jeopardy won tegen de beste menselijke spelers aller tijden) en ook AlphaGo (die in 2016 de Roger Federer van het go-spel, Lee Sedol, versloeg) verbruiken megawatts aan energie. Van Gerven: „Stel dat neurale netwerken het juiste model zijn om mensachtige kunstmatige intelligentie mee te bouwen. Dan is het onhandig om dat te doen op computers die informatie achter elkaar verwerken, zoals nu gebeurt, in plaats van parallel, zoals het brein dat doet. Wat je wilt is het inpakken van parallele informatieverwerking in de hardware zelf. Het vakgebied van de neuromorphic computing houdt zich daarmee bezig. Maar zowel voor deze hardware-kant als voor de software-kant van neurale netwerken geldt: het gat tussen de systemen die we nu hebben en wat ons brein doet, is gigantisch groot. Beweringen dat dat niet zo is, moet je met een korrel zout nemen.”

Iris van Rooij denkt dat als de wetenschap van de kunstmatige intelligentie echt mensachtige intelligentie in een machine wil bouwen, ze er niet aan ontkomt om ook te kijken naar resultaten uit de psychologie, de cognitiewetenschappen, de neurowetenschappen, de filosofie en de theoretische informatica. „Anders kijk je in het donker. Dan is het alsof je een vliegtuig gaat bouwen zonder kennis van de aerodynamica.” Evenmin als Josh Tenenbaum en Marcel van Gerven ziet ze echter een fundamentele beperking aan het bouwen van machines met een mensachtige intelligentie.

„Ja, ook met empathie en bewustzijn,” besluit Van Gerven. „Het enige andere alternatief is dat je een magisch ingrediënt veronderstelt, en ik zou niet weten welk.”

Inlevingsvermogen Peuter opent kast



Beelden uit een filmpje van psychologen Felix Warneken en Michael Tomasello, die altruïsme bij jonge kinderen onderzochten. Cognitiewetenschapper Josh Tenenbaum illustreert met deze film de beperkingen van kunstmatige intelligentie.