

ChatGPT en de mens

Emanuel Rutten

Introductie

Vrijwel direct nadat ChatGPT eind november vorig jaar door het Amerikaanse bedrijf OpenAI geïntroduceerd werd, stond het overal in de wereld volop in de belangstelling. Deze interesse houdt nog altijd aan. De droom van kunstmatige intelligentie lijkt eindelijk waargemaakt. Of toch niet? Waartoe zal ChatGPT uiteindelijk in staat zijn en wat kan het nu al? Is wat deze software doet te begrijpen als denken? Moeten we zeggen dat ChatGPT komt tot begrip en oordeelsvorming?

Dit zijn vragen die velen terecht bezighoudt. Halverwege de negentiende eeuw heeft Charles Darwin met zijn monumentale boek *On the Origin of Species* de mensheid laten zien dat het ontstaan van complexe levende organismen goed kan worden verklaard vanuit slechts enkele eenvoudige evolutionaire selectieregels. Bijna twee eeuwen later laat ChatGPT op zijn beurt de mensheid zien dat het voortbrengen van complexe betekenisvolle taaluitingen goed kan worden verklaard vanuit slechts enkele eenvoudige statistische voorspellingsregels. Is ChatGPT daarmee de Darwin van betekenis?

In elk geval is Darwins evolutionaire verklaring niet van toepassing op het voortbrengen van betekenisvolle teksten en gesprekken. Het genereren van organismen op grond van een of meerdere selectieregels is namelijk iets anders dan het genereren van betekenissen op grond van een of meerdere voorspellingsregels. Maar hoewel ChatGPT heeft laten zien dat het met zijn statistische voorspellingsregels in tegenstelling tot Darwins evolutionaire selectieregels gecompliceerde betekenisvolle dialogen kan produceren, zal ik in wat volgt betogen dat uiteindelijk ook ChatGPT er niet in slaagt om werkelijk tot begrip of het verstaan van betekenis te komen.

Hoe werkt ChatGPT? ChatGPT traint zichzelf uitsluitend op wat wij gedacht, gezegd en geschreven hebben en in de afgelopen decennia op het internet beschikbaar hebben gemaakt. Het is daarom een spiegel voor de mens. We praten voortdurend met onszelf en ons verleden. ChatGPT zoekt patronen in wat mensen voortgebracht hebben en geeft er allerlei combinaties van terug. Het maakt dan ook nooit een werkelijk originele creatieve sprong. Het *opent* niets. En daarom denkt het niet. De menselijke geest kan verbanden leggen tussen ogenschijnlijk volstrekt niet gerelateerde gegevens om zo uit het al bekende te springen en nieuwe ideeën te creëren die voorheen niet bestonden. ChatGPT kan dit niet. Het kan louter samenstellingen

maken van wat wij al gedacht hebben. Het loopt anders gezegd eindeloos achter ons denken aan met louter permutaties van het door ons reeds gedachte. Dat is geen denken.

Geen waarheid en vrijheid

ChatGPT mist bovendien twee cruciale ingrediënten voor werkelijk denken. In de eerste plaats heeft ChatGPT geen besef van waar en onwaar. Want het berekent en retourneert slechts statistisch waarschijnlijke uitbreidingen van aangeboden woorden. En omdat een conceptie van waar en onwaar een noodzakelijke voorwaarde is voor begrip, heeft ChatGPT geen werkelijk begrip. Er is daarom helemaal geen sprake van het verstaan van betekenis. Precies omdat ChatGPT een conceptie mist van wat waar en onwaar überhaupt betekenen, kan het niet beoordelen of iets waar of onwaar is. Het mist een representatie of een model van de wereld. Kortom, ChatGPT heeft geen waarheidsbesef en dáárom denkt het niet. Het is dus niet in staat tot werkelijke oordeelsvorming. Een computerprogramma dat betekenisvolle conversaties nabootst door patronen in grote menselijke datasets te ontsluiten, zonder daadwerkelijk te begrijpen wat de concepten in de dataset betekenen, moet niet als begrijpend of denkend worden beschouwd. Dat ChatGPT allerlei complexe puzzels kan oplossen doet hieraan niets af. Zelfs als het op een dag een beroemd onopgelost wiskundig probleem zou oplossen, zou ik exact hetzelfde zeggen als toen de wereldkampioen schaken voor het eerst werd verslagen door een computerprogramma. Het zou algoritmisch buitengewoon vernuftig zijn, maar het zou geen denken zijn.

In de tweede plaats is voor echt denken, voor werkelijk conceptueel begrip, innerlijke vrije subjectiviteit of een vrij bewustzijn nodig is. Een vrij subject verschilt van nature echter radicaal van elk gesloten formeel regelsysteem. ChatGPT is als formeel systeem gesloten. Alleen een vrije bewuste geest is wezenlijk open en in staat om elk gesloten formalisme te overstijgen. Dát is wat eigenlijk denken is. Eigenlijk denken valt niet met een enkel gesloten formalisme samen.

Natuurlijk maakt ChatGPT indruk. Op het eerste gezicht lijkt het erop dat ChatGPT in staat is om begrip en betekenis te herleiden tot structuur en *syntax*. Het lijkt daarmee een kloof te overbruggen die onoverbrugbaar leek. Maar omdat wat ChatGPT doet geen werkelijk denken is, is van een herleiding van begrip tot syntax geen sprake. De sfeer van begrip en betekenis betreft immers het denken. De indruk die ChatGPT aanvankelijk maakt, wordt dan ook getemperd zodra we iets preciezer begrijpen op basis van welke principes het werkt. ChatGPT voorspelt steeds hoe een ingevoerde reeks van woorden verlengd zou kunnen worden door statistische voorspellingsmodellen te gebruiken die het algoritmisch genereert op basis van een analyse van grote verzamelingen van op internet beschikbare gegevens. Gelijksortige

reeksen van woorden worden zo in kaart gebracht en productief gemaakt zodra ChatGPT als invoer een rij van woorden van de gebruiker ontvangt en hierop moet reageren. ChatGPR ziet dus geen verbanden die datgene overstijgen waartoe de gegevensverzamelingen statistisch aanleiding geven. Onvoorziene verbanden buiten de beschikbare teksten ontgaan ChatGPT.

Geen paradigmawisselingen

Precies omdat ChatGPT zoals gezegd slechts eindeloos variaties op het al bestaande teruggeeft, is het, in tegenstelling tot de mens, niet in staat tot vakwetenschappelijke of filosofische paradigmawisselingen. Dat ChatGPT niet denkt, dat het buiten wat wij zelf denken nooit een opening maakt naar iets werkelijk nieuws, heeft anders gezegd tot gevolg dat het geen nieuwe concepten kan scheppen buiten de reeds bekende concepten. ChatGPT breidt daarom onze conceptuele ontologie niet uit. En zonder een dergelijke ontologische uitbreiding is het onmogelijk om tot een nieuw paradigma te komen.

Uiteraard is veel van wat mensen origineel en creatief noemen niet werkelijk creatief en origineel in hierboven bedoelde zin. Werkelijke creativiteit is zeldzaam. Neem bijvoorbeeld het cartesiaanse wereldbeeld dat opkwam aan het begin van de zeventiende eeuw en radicaal brak met het klassieke wereldbeeld van Aristoteles. Het cartesianisme is vergeleken met het aristotelische wereldbeeld een voorbeeld van een werkelijk originele creatieve sprong. En hetzelfde geldt voor de natuurkunde van Einstein ten opzichte van die van Newton. Hier worden totaal nieuwe conceptuele ruimten ontsloten. Tot dit soort sprongen naar nieuwe ideeënwerelden is ChatGPT niet in staat. De sprong van Einstein is en blijft dan ook een volstrekt oorspronkelijke sprong van de menselijke geest die op geen enkele wijze gemaakt had kunnen worden door patroonherkenning in teksten van Newton, Kant en andere voorlopers. Het gaat om een nieuw paradigma, een nieuw *episteme* om met Foucault te spreken, en dat lukt niet met alleen maar patroonherkenning in het bestaande. ChatGPT zou in een klassiek aristotelische wereld op basis van aristotelische begrippen slechts eindeloos variaties op aristotelische teksten herhalen en dus nooit een origineel cartesiaans denken ontsluiten en laten verschijnen. Hoewel ChatGPT dus op grond van bestaande gegevens binnen bestaande domeinen nieuwe verbanden kan vinden tussen bestaande concepten, kan het nimmer nieuwe conceptuele domeinen creëren. Gebruikmakend van door ons al vaak gebruikte stijlen blijft het slechts binnen de door ons al veel besproken onderwerpen. Hoewel de mens veelal dezelfde techniek toepast, doet de mens dat niet altijd omdat zoals gezegd de menselijke geest open is en precies daarom alle gesloten regelgeleide formalismen overstijgt.

Een teloorgang van de menselijke existentie?

Door zeer intensief met ChatGPT aan de slag te gaan en ervan in de ban te raken, loopt de

mensheid echter het risico zich uitsluitend te gaan richten op wat ze reeds gedacht heeft. Er zullen veel minder originele creatieve spongen naar het onbekende plaatsvinden. Worden er straks nog wel nieuwe conceptuele werelden geopend? Dit is een existentiële vraag. Door de eeuwen heen is de wereld steeds op een andere wijze aan de mens verschenen. De manier waarop de dingen in de klassieke oudheid in de openbaarheid traden, verschilt wezenlijk van de wijze waarop de dingen in de middeleeuwen in het licht traden of van de manier waarop de dingen in de moderne tijd verschenen en zo blijk gaven van hun aanwezigheid. Zal door een dominantie van ChatGPT en soortgelijke technologieën de wereld mogelijk nooit meer op een andere wijze aan de mens verschijnen? Zullen de dingen nooit meer op een geheel nieuwe manier voor ons ontsloten worden en tevoorschijn treden? Deze vragen raken aan de aard of het wezen van de mens. Het risico dat wij onze geestelijke openheid naar de wereld en de toekomst verliezen en zo onze verwonderende ontvankelijke bestaanservaring vervalt tot louter monotone voorspelbare afgeslotenheid is niet denkbeeldig. Als dit gebeurt, verliest de menselijke existentie enorm aan zin en betekenis.

Vanuit Heideggers ontologische denken over de saamhorigheid van mens, taal en wereld kan ChatGPT op twee manieren begrepen worden als een ontoelaatbare *ontische* verschraling. In de eerste plaats is het grote taalmodel (*large language model*) van ChatGPT gebaseerd op de veronderstelling dat het analyseren en algoritmisch verwerken van grote verzamelingen van *taaluitingen* een veelbelovende weg is om gedachten voort te brengen. Deze veronderstelling is een ontische versmalling van Heideggers ontologische these dat denken en taal ten diepste existentieel op elkaar betrokken zijn. Taal is geen neutraal voertuig of medium voor het overbrengen van inhouden, maar veeleer de fundamentele manier waarop het denken zich ontvouwt en begrip van de wereld mogelijk maakt. De taal opent een geestelijke ruimte waarin en van waaruit het denken betekenisvol betrokken kan zijn op de wereld. Genoemde veronderstelling van ChatGPT's grote taalmodel is preciezer gezegd een rationalistische vernauwing van Heideggers existentiële these omdat volgens deze these authentiek denken neerkomt op het aandachtig luisteren naar de stem van de taal oftewel naar wat de taal ons ten diepste te zeggen heeft en ons te zeggen geeft – en zo voor ons betekenisvol onthult.

In de tweede plaats is ChatGPT's grote taalmodel uitsluitend geïnteresseerd in een neurale statistische analyse van *vroegere* datasets. Dit is eveneens een ontische verschraling van een belangrijke ontologische these van Heidegger, namelijk de existentiële these dat ons verleden diep verweven is met ons zelfbegrip als historische wezens, zodat echt denken geworteld is in de herinnering en dus in wat de overlevering of traditie ons toont. Echt denken is volgens deze tweede these openstaan naar de toekomst vanuit een ontvankelijke verwonderende bezinning op wat geweest is. Wat ChatGPT doet kan dus op minimaal twee manieren worden

beschouwd als een rationalistische vernauwing van Heideggers bezinnende existentiële denken over wat de mens en menselijk denken daadwerkelijk is. Dit zou ons op z'n minst moeten waarschuwen. ChatGPT doet geen recht aan de wijze waarop wij als denkende wezens in de wereld zijn. Het ontketent daarom uiteindelijk het wezen van de mens.

Inderdaad, de menselijke geest is niet gesloten. De menselijke geest staat open voor vele nieuwe en nog onvoorziene mogelijkheden, terwijl ChatGPT zoals besproken slechts op eerdere patronen let om vervolgens alleen maar vergelijkbare patronen te creëren. Een onbeheerde adoptie van ChatGPT met zijn grote ontische taalmodel kan daarom op termijn mogelijk leiden tot een afsluiting van de menselijke geest en daarmee tot het verdwijnen van de menselijke existentie. We stuiten zo op een mogelijk gevolg dat uiteraard van een geheel andere orde is dan de inmiddels overal besproken meer voor de hand liggende gevolgen van ChatGPT, zoals die voor de werkgelegenheid, de fraudebestendigheid van het onderwijs en het verdienmodel van bedrijven. Niets minder dan de mens zelf staat op het spel. Wat nodig is, is daarom een meer fundamentele doordenking van de hier beknopt geschatste risico's van een grenzeloos gebruik van ChatGPT voor onze leefwereld en de menselijke bestaanservaring.