Libertarische vrije wil als conditie voor rationeel delibereren en de implicaties hiervan voor de cognitieve mogelijkheden van kunstmatige intelligenties

Emanuel Rutten

VU Amsterdam

Vrijwel direct nadat ik gevraagd werd om een bijdrage te leveren aan de afscheidsbundel van René schoot mij zijn artikel *A No-Miracle Argument for Libertarianism* te binnen dat vele jaren geleden is besproken tijdens een onderzoeksbijeenkomst en mij sindsdien altijd is bijgebleven. In dit artikel geeft René in twee stappen een argument voor de bewering dat mensen een libertarische vrije wil hebben. Zijn eerste stap vertrekt vanuit het gegeven dat mensen regelmatig met succes intelligente complexe taken voltooien. Gebruikmakend van aanvullende premissen leidt hij hieruit af dat mensen een vrije wil hebben. Ik geef hieronder Renés eerste stap op p. 15 van zijn artikel schematisch weer.

1. We have SUCCESS [at complex tasks].
2. For SUCCESS, a significant number of true beliefs relevant to SUCCESS are required.
3. Hence, we have a significant number of true beliefs relevant to SUCCESS. <from (1) and (2)>
4. In order to acquire that significant number of true beliefs relevant to SUCCESS, deliberation is required.
5. We cannot deliberate unless we believe we have free will.
6. In order to acquire that significant number of true beliefs relevant to SUCCESS, we must believe we have free will. <from (4) and (5)>
7. If belief B is required for SUCCESS, and we have SUCCESS, then, defeasibly, the belief B is true.
8. Hence, if the belief that we have free will is required for the acquisition of true beliefs relevant to SUCCESS, and we have SUCCESS, then, defeasibly, the belief that we have free will is true. <from (7)>
9. Hence, defeasibly, our belief in free will is true. <from (6), (3), and (8)>

De tweede stap vertrekt vanuit het bestaan van de vrije wil. Door een beroep op andere additionele premissen te doen, beargumenteert René dat de vrije wil incompatibel is met determinisme, zodat uiteindelijk de conclusie volgt dat mensen een libertarische vrije wil hebben. De tweede stap van het argument van René, zoals vermeld op p. 18 van zijn artikel, volgt hieronder. Hoewel ik mij niet meer kan herinneren wat ik tijdens genoemde onderzoeksgroep inbracht, wil ik hier, om René behulpzaam te zijn, opmerken dat ik hieronder enkele kleine schrijffoutjes in zijn weergave op p. 18 hersteld heb.

10. We have free will.
11. If we have free will, either compatibilism or libertarianism is true.
12. Hence, either compatibilism or libertarianism is true. <from (10) and (11)>
13. If compatibilism is true, SUCCESS is a miracle.
14. SUCCESS is not a miracle.
15. Hence, compatibilism is false. <from (13) and (14)>
16. Hence, libertarianism is true. <from (12) and (15)>

Zijn argument is fascinerend en ik wilde er voor zijn afscheidsbundel graag over schrijven. René liet mij onlangs echter weten dat zijn artikel op de plank is blijven liggen omdat andere projecten naar zijn mening voorrang verdienden. Dit bracht mij toch niet van mijn voornemen af. Uiteraard zonder prijs te geven dat er een afscheidsbundel aankwam, vroeg en kreeg ik van hem toestemming om zijn artikel in te brengen in een stuk over de cognitieve reikwijdte van kunstmatige intelligentie. Dat stuk

is deze bijdrage geworden. In wat volgt beargumenteer ik eerst dat zijn eerste stap al tot de conclusie leidt dat mensen een libertarische vrije wil hebben. René is dus te bescheiden in zijn opvatting dat de beoogde conclusie pas in zijn tweede stap wordt bereikt. Door zijn argumentatie in de eerste stap op de voet te volgen, geef ik aan waar en hoe deze stap reeds tot de beoogde conclusie leidt. Onderweg doe ik ook enkele suggesties om zijn eerste stap argumentatief nog wat scherper te maken, waarmee René, als hij zich in deze suggesties kan vinden, in een nieuwe versie van zijn artikel zijn voordeel kan doen. Daarna laat ik zien dat we aan de eerste stap van René bepaalde inzichten kunnen ontlenen die helpen bij het nadenken over de cognitieve grenzen van kunstmatige intelligenties. Zo kunnen we ons bijvoorbeeld afvragen of kunstmatige intelligenties kunnen denken. Deze vraag is wijsgerig relevant omdat het samenhangt met vragen omtrent de aard van bewustzijn en intentionaliteit, en omdat het ethische implicaties heeft voor de morele status en rechten van kunstmatige intelligenties. Ik betoog eerst zonder gebruik te maken van het argument van René dat er goede redenen zijn om te beweren dat kunstmatige intelligenties niet kunnen denken en ik laat zien tot welke cognitieve beperkingen voor deze intelligenties dit leidt. Vervolgens beargumenteer ik op grond van aan de eerste stap van René ontleende inzichten dat kunstmatige intelligenties niet kunnen denken. Het zal blijken dat deze inzichten onafhankelijk zijn van genoemde goede redenen. De eerste stap van René versterkt daarom de conclusie dat kunstmatige intelligenties niet kunnen denken. Zijn argument is dan ook niet alleen van belang voor het aloude debat over het bestaan van de vrije wil, maar blijkt eveneens vruchtbaar voor het debat over de omvang van de cognitieve capaciteiten van kunstmatige intelligenties.

De conclusie van de eerste stap is dat mensen een vrije wil hebben. René begrijpt hier vrije wil aldus: ‘A person has acted out of free will, when she faced a choice between alternative but incompatible courses of action, each of which is ‘up to her’, and she selected one of those courses, and pursued it.’ Aanhangers van de libertarische vrije wil bedoelen met de uitdrukking ‘up to her’ dat het niet causaal gedetermineerd was welke actie zij koos en dat ze dus onder dezelfde omstandigheden anders had kunnen kiezen dan ze deed. Deze libertarische betekenis van ‘up to her’ veronderstelt René hier niet omdat hij in de eerste stap niet argumenteert voor een libertarische vrije wil. Dat doet René pas in zijn tweede stap. Compatibilisten bedoelen met ‘up to her’ iets anders, namelijk dat haar keuze overeenstemt met haar voorkeuren en niet ontstond door externe dwang of externe beperkingen. Afwezigheid van externe belemmeringen maakt haar keuze vrij. En ze zou een andere keuze gemaakt hebben als ze andere voorkeuren gehad zou hebben en deze voor hun realisatie een andere keuze zouden hebben vereist. Deze compatibilistische betekenis van ‘up to her’ veronderstelt René hier ook niet omdat hij in zijn tweede stap concludeert dat compatibilisme onwaar is. René heeft dus voor zijn eerste stap een betekenis van ‘up to her’ op het oog die consistent is met zowel een libertarische als compatibilistische vrije wil. René expliciteert deze betekenis op deze plaats in zijn tekst echter niet.

De eerste stap van het argument van René is gebaseerd op vijf premissen. De eerste premisse is zoals gezegd het feit dat mensen herhaaldelijk succesvol complexe taken volbrengen, zoals het bouwen van schepen en het bevaren van de oceanen. Een complexe taak bestaat uit causaal geschakelde en aan externe invloeden onderhevige acties die op grond van nieuwe informatie aangepast kunnen worden. Volgens René moeten deterministen accepteren dat succes metafysisch contingent is omdat zo’n taak faalt in tenminste één mogelijke wereld met andere natuurwetten en begincondities. Een determinist kan tegenwerpen dat ook de natuurwetten en de begincondities deterministisch bepaald zijn. Maar dit levert geen probleem voor René op omdat de eerste premisse geen beroep op contingentie doet.

De tweede premisse stelt dat het succesvol uitvoeren van een complexe taak een significant aantal ware voor succes relevante overtuigingen vereist. Zonder voor de taak relevante ware overtuigingen zal inderdaad geen succes behaald worden. Volgens de derde premisse is deliberatie vereist om tot genoemde overtuigingen te komen. Er dient immers voortdurend onderzoek plaats te vinden om

dergelijke overtuigingen te verkrijgen en onderzoek vereist deliberatie. René begrijpt deliberatie als een proces waarbij gekozen moet worden uit verschillende acties. Deliberatie eindigt in een beslissing om de gekozen actie uit te voeren. En dit is onbetwistbaar onmisbaar voor het doen van onderzoek.

De vierde premisse luidt dat deliberatie vereist dat degene die delibereert gelooft dat hij of zij een vrije wil heeft. Volgens René is delibereren bezig zijn met het maken van een keuze uit alternatieve acties en 'he cannot do that unless he believes that it is in his power to do each of these. That is to say, he cannot deliberate unless he believes he has free will.' De betekenis van 'in his power' laat René hier onbesproken. Het moet wederom gaan om een betekenis die consistent is met zowel een libertarische als compatibilistische vrije wil omdat de conclusie van zijn eerste stap dat mensen een vrije wil hebben consistent is met een libertarische en compatibilistische vrije wil. Pas in zijn tweede stap beargumenteert hij immers dat het gaat om een libertarische en niet compatibilistische vrije wil.

Wie delibereert moet volgens de vierde premisse dus geloven een vrije wil te hebben. Volgens René heeft Derk Pereboom deze premisse betwist door te beargumenteren dat ook iemand die gelooft dat al zijn acties causaal gedetermineerd zijn rationeel kan delibereren. De claim van Pereboom weerlegt de vierde premisse van René echter niet. Geloof in de waarheid van determinisme sluit geloof in de vrije wil namelijk niet uit. En dit omdat het in de eerste stap vooralsnog om een compatibilistische vrije wil mag gaan. René weerlegt compatibilisme immers pas in de tweede stap van zijn argument.

Dat René toch meent dat Perebooms claim in tegenspraak is met de vierde premisse laat dus zien dat René vanaf in elk geval dit moment van zijn eerste stap in feite werkt met de libertarische betekenis van vrije wil. Dit is goed nieuws, want als deze stap argumentatief slaagt, dan wordt al aan het eind ervan de conclusie bereikt dat mensen een libertarische vrije wil hebben.

Hoe slaat René de aanval van Pereboom af? Pereboom geeft twee condities die volgens hem elk consistent zijn met geloof in determinisme en waarvan de conjunctie een voldoende voorwaarde is voor rationeel delibereren. De eerste conditie stelt dat wie rationeel delibereert geen actie overweegt waarvan zij zeker weet deze te zullen kiezen óf waarvan zij zeker weet deze niet te zullen kiezen óf waarvan zij gelooft dat deze in tegenspraak is met wat zij gelooft. De tweede conditie stelt dat wie rationeel delibereert gelooft dat zij normaal gesproken zal doen wat ze gekozen heeft. Beide condities zijn inderdaad consistent met geloof in determinisme. Als eveneens de conjunctie een voldoende voorwaarde is voor rationeel delibereren, dan is de vierde premisse van René weerlegd.

René meent echter het volgende tegenvoorbeeld te kunnen geven. Als zijn favoriete voetbalteam de eerstvolgende wedstrijd wint, dan wil René bij de wedstrijd daarna aanwezig zijn en daartoe kan hij de trein of het vliegtuig nemen. René merkt terecht op dat hier aan beide condities is voldaan. Maar toch, zo stelt René, is hij nog niet aan het delibereren, zodat dit voorbeeld volgens hem laat zien dat de conjunctie van beide condities geen voldoende voorwaarde is voor deliberatie.

Dit volgt echter niet. De condities van Pereboom zijn immers condities voor *rationeel* delibereren. Ze zijn daarom alleen van toepassing op situaties waarin sprake is van deliberatie en we ons afvragen of deze deliberatie al dan niet rationeel is. Pereboom kan dus tegenwerpen dat hier geen sprake is van een tegenvoorbeeld omdat er überhaupt nog niet gedelibereerd wordt en dus de vraag of er sprake is van rationeel delibereren niet eens kan opkomen. Vergelijk dit met het volgende. Wie van mening is dat afwezigheid van geloof in God geen voldoende voorwaarde is voor het zijn van atheïst, kan als tegenvoorbeeld niet inbrengen dat voetbalstations niet in God geloven en toch geen atheïst zijn. Want wie meent dat afwezigheid van geloof in God een voldoende voorwaarde is voor atheïsme, meent op niet controversiële wijze dat deze voorwaarde alleen van toepassing is op situaties waarin reeds sprake is van een persoon waarvan wij ons redelijkerwijs kunnen afvragen of deze wel of geen atheïst is. Hetzelfde geldt voor Perebooms condities en de vraag of deze een voldoende voorwaarde

zijn voor *rationeel* delibereren. Hierbij wordt reeds uitgegaan van een situatie van deliberatie. En dit is mogelijk één van de niet controversiële aanvullende condities van Pereboom waarop René wijst.

Het voorbeeld van René ondermijnt Perebooms onderbouwing voor zijn claim en daarmee de claim zelf dus niet. Voor de eerste stap heeft dit echter geen argumentatieve gevolgen omdat René naast genoemd voorbeeld ook een rechtstreekse weerlegging van Perebooms claim geeft. Zijn weerlegging verloopt in zijn woorden verkort weergegeven als volgt. Waarom oftewel met welk doel delibereren deterministen? Wat betekent het voor een determinist om uit te zoeken wat te doen en te kiezen uit verschillende acties indien er volgens deterministen altijd maar één actie fysisch mogelijk is en er van een werkelijke keuze dus nooit sprake is? Delibereren is zinloos voor een determinist. Het dient geen enkel doel. Wat René hier feitelijk laat zien is dat geloof in determinisme weliswaar kan samengaan met delibereren, maar niet kan samengaan met *rationeel* delibereren. Wie *rationeel* delibereert kan niet anders dan geloven dat er werkelijk iets te kiezen valt en dat het kiezen dus geen lege huls is. Een determinist kan dus alleen maar *irrationeel* delibereren. Perebooms claim is hiermee weerlegd, zodat de vierde premisse is veiliggesteld. Zowel de derde als vierde premisse moeten vanaf nu echter wel preciezer geformuleerd worden. De derde premisse moet luiden dat *rationele* deliberatie nodig is om genoeg relevante ware overtuigingen te verkrijgen voor het succesvol voltooien van een intelligente complexe taak en de vierde premisse dat *rationele* deliberatie bij de beraadslager geloof in de vrijheid van zijn wil vereist. Deze aanpassingen kunnen echter argumentatief probleemloos gemaakt worden.

René eindigt zijn weerlegging overigens met de volgende opmerking: ‘Under these conditions, there isn’t really any such thing as ‘deciding between incompatible actions’ […]. For that requires that each of the alternatives are ‘up to the agent’, which (if determinism is true) they are not.’ De waarheid van ‘Being up to the agent’ impliceert dus de onwaarheid van determinisme. René werkt inmiddels in zijn eerste stap dus inderdaad met de libertarische betekenis van ‘up to the agent’ en dus van de vrije wil.

De vijfde premisse kan als volgt geformuleerd worden. Als geloof in propositie P vereist is voor het regelmatig succesvol voltooien van complexe taken, en als we dergelijke taken regelmatig succesvol voltooien, dan is, in afwezigheid van redenen om het tegendeel te denken, P waar. René bespreekt in zijn onderbouwing van deze premisse een verwant principe van David Enoch: Als geloof in propositie P onmisbaar is voor het uitvoeren van een respectabel project dat bovendien onmisbaar is voor de mens (omdat wij onszelf zouden ondermijnen als we ermee zouden stoppen), dan is, in afwezigheid van redenen om het tegendeel te denken, geloof in P respectabel en dus rationeel gerechtvaardigd. René meent dat Enoch’s principe niets bijdraagt aan de onderbouwing van zijn vijfde premisse. Ook hier is René te bescheiden. Het draagt daaraan wél bij. Beschouw de volgende these: Als geloof in P vereist is om stand van zaken A te actualiseren, en als we A actualiseren, dan is, in afwezigheid van redenen om het tegendeel te denken, P waar. De vijfde premisse is een instantie van deze these. De empirische ondersteuning die René voor zijn vijfde premisse biedt, komt in feite neer op empirische ondersteuning voor deze these. Neem voor A maar epistemisch succes in de wetenschap en voor P de propositie dat de wereld ordelijk is of dat ons verstand grotendeels betrouwbaar is. Op grond van Enoch’s overwegingen kunnen we de these waarvan Renés premisse een instantie is, vervangen door een minder vergaande en daarom plausibelere these: Als geloof in P vereist is om stand van zaken A te actualiseren, en als A voor ons onmisbaar is omdat wij zonder A onszelf zouden ondermijnen, en als we A actualiseren, dan is, in afwezigheid van redenen om het tegendeel te denken, P waar. De vijfde premisse is ook een instantie van deze plausibelere these. En de plausibelere these wordt ook empirisch ondersteund door Renés empirische ondersteuning. Neem voor A en P maar hetzelfde als hierboven. Enouch’s principe versterkt dus wel degelijk Renés onderbouwing van zijn vijfde premisse. Door gebruik te maken van de kwalificatie ‘in afwezigheid van redenen om het tegendeel te denken’ falen bovendien, zoals René ook laat zien, bepaalde tegenvoorbeelden die anders zouden slagen.

Nu alle premissen besproken zijn, kan Renés eerste stap wederom formeel weergegeven worden. Hoewel ik mij zoals eerder gezegd niet meer kan herinneren wat ik toen tijdens de onderzoeksgroep inbracht, wil ik hier, wederom om René behulpzaam te zijn, opmerken dat in zijn formalisering op p. 15 stap (3) overbodig is, (8) volgt uit (7) *en (2)*, en (9) volgt uit (6), *(1)* en (8). Hieronder volgt een vertaalde en bijgewerkte formele weergave van Renés eerste stap (i.e., *a* t/m *h*). Hierbij wordt 'succes' begrepen als het regelmatig succesvol voltooien van intelligente complexe taken.

a. Mensen hebben succes *(eerste premisse)*,
b. Een significant aantal ware relevante overtuigingen is vereist voor succes *(tweede premisse)*,
c. Rationeel delibereren is vereist voor het verkrijgen van zulke overtuigingen *(derde premisse)*,
d. Geloof in het hebben van een vrije wil is vereist voor rationeel delibereren *(vierde premisse)*,
e. Geloof in het hebben van een vrije wil is vereist voor succes *(uit b, c en d)*,
f. Als geloof in propositie P vereist is voor succes, en we hebben succes, dan is, in afwezigheid van redenen om het tegendeel te denken, P waar. *(vijfde premisse)*,
g. Als geloof in het hebben van een vrije wil vereist is voor succes, en we hebben succes, dan is, in afwezigheid van redenen om het tegendeel te denken, 'Onze wil is vrij' waar *(uit f)*,
h. Mensen hebben, in afwezigheid van redenen voor het tegendeel, een vrije wil *(uit e, a, g)*.

Het gaat hier zoals besproken om een libertarische vrije wil, zodat redelijkerwijs volgt dat mensen een libertarische vrije wil hebben. Dat direct al in Renés eerste stap de conclusie bereikt wordt dat wij een libertarische en geen compatibilistische vrij wil hebben, hoeft ons overigens niet te verbazen. In het algemeen argumenteren voor het bestaan van de vrije wil heeft namelijk niet zoveel zin. We zullen eerst expliciet moeten aangeven wat we met de vrije wil bedoelen voordat we een zinvol argument kunnen geven voor het bestaan ervan. En omdat de compatibilistische vrije wil dermate afwijkt van de libertarische vrije wil, zal dus direct al in de eerste stap aangegeven moeten worden voor welke vrije wil eigenlijk geargumenteerd wordt. Vergelijk dit met het argumenteren voor het bestaan van morele waarden. Het maakt nogal uit of het gaat om objectieve van de mens onafhankelijke waarden of om van de mens afhankelijke intersubjectieve sociale conventies. Dit zal eerst verduidelijkt moeten worden voordat een zinvol argument voor het bestaan van morele waarden gegeven kan worden.

Tot zover de bespreking van de succesvolle eerste stap van René. René beperkt zich tot de mens als cognitieve actor. Er is echter geen reden om te denken dat zijn eerste stap niet van toepassing is op andere natuurlijke cognitieve actoren. Bovendien gaf ik in de inleiding aan dat aan Renés eerste stap ontleende inzichten ook relevant zijn voor ons nadenken over het cognitieve bereik van kunstmatige cognitieve actoren oftewel kunstmatige intelligenties. Zoals eveneens in de inleiding aangekondigd, betoog ik, om dit te laten zien, hieronder eerst, op grond van overwegingen die onafhankelijk zijn van Renés argument, dat kunstmatige intelligenties niet in staat zijn tot daadwerkelijk denken. Hiertoe richt ik mij op ChatGPT als exemplarisch of paradigmatisch voorbeeld van dergelijke intelligenties.

***

Waartoe zal ChatGPT uiteindelijk in staat zijn en wat kan het nu al? Is wat het doet te begrijpen als denken? Moeten we zeggen dat ChatGPT komt tot begrip en oordeelsvorming? Dit zijn vragen die velen terecht bezighoudt. In de negentiende eeuw heeft Darwin ons laten zien dat het ontstaan van complexe organismen kan worden verklaard vanuit slechts enkele zeer eenvoudige evolutionaire selectieregels. ChatGPT laat ons twee eeuwen later zien dat het voortbrengen van allerlei complexe betekenissen kan worden verklaard vanuit slechts enkele zeer eenvoudige statistische zoekregels. Is ChatGPT daarom de Darwin van betekenis? In elk geval is Darwins evolutionaire verklaring niet van toepassing op het voortbrengen van betekenissen. Het genereren van organismen op grond van enkele selectieregels is iets anders dan het op grond van enkele zoekregels genereren van

betekenissen. Maar hoewel ChatGPT met zijn zoekregels in tegenstelling tot Darwins selectieregels betekenissen kan genereren, slaagt ChatGPT er niet in om werkelijk tot begrip en betekenis te komen.

Dit volgt uit een reflectie op hoe ChatGPT werkt. Het traint zichzelf uitsluitend op wat wij gedacht, gezegd en geschreven hebben. ChatGPT is daarom een spiegel voor de mens. We praten voortdurend met onszelf en ons verleden. ChatGPT zoekt alleen 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. Het denkt niet. De menselijke geest kan oorspronkelijke verbanden leggen tussen ogenschijnlijk volstrekt ongerelateerde gegevens om zo uit het al bekende te springen en geheel nieuwe ideeën te creëren die voorheen niet bestonden. ChatGPT kan dit niet. Het kan louter samenstellingen maken van wat wij als mensen al gedacht hebben. En dat is geen denken. Bovendien heeft ChatGPT geen enkel besef van waar en onwaar. 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 dus geen sprake van echte betekenis. Precies omdat ChatGPT een concept mist van wat waar en onwaar überhaupt betekenen, kan het helemaal niet beoordelen of iets waar of onwaar is. ChatGPT heeft geen waarheidsbesef en dáárom denkt het niet. Het is 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 het allerlei puzzels kan oplossen doet hier niets aan 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. Voor denken, voor werkelijk conceptueel begrip is innerlijke vrije subjectiviteit of een vrij bewustzijn nodig dat van nature radicaal verschilt van elk gesloten formeel regelsysteem. Alleen een vrije bewuste geest is wezenlijk open en in staat om elk gesloten formalisme te overstijgen. Dat is wat eigenlijk denken is. Eigenlijk denken valt met geen enkel gesloten formalisme samen.

Natuurlijk maakt ChatGPT veel indruk. Op het eerste gezicht lijkt het erop dat ChatGPT in staat is om begrip en betekenis te reduceren 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 reductie van begrip tot syntax geen sprake. De sfeer van begrip en betekenis betreft immers het denken. De grote indruk die ChatGPT aanvankelijk maakt, wordt getemperd zodra we begrijpen op basis van welke principes het werkt. Het voorspelt statistisch hoe een ingevoerde reeks van woorden verlengd zou moeten worden door grote verzamelingen van teksten te doorzoeken naar gelijksoortige reeksen. Het ziet dus geen verbanden die datgene overstijgen waartoe de gegevensverzamelingen statistisch aanleiding geven. Onvoorziene verbanden buiten de beschikbare teksten ontgaan ChatGPT. In tegenstelling tot de mens is ChatGPT dan ook niet in staat tot paradigmawisselingen. Het geeft slechts variaties op het al bestaande terug. Precies omdat ChatGPT niet denkt, zal het buiten wat wij zelf denken nooit een opening maken naar iets werkelijk nieuws. Het schept geen nieuwe concepten buiten de gegeven bestaande concepten. Het breidt onze conceptuele ontologie niet uit. ChatGPT loopt eindeloos achter ons denken aan met louter permutaties van het door ons reeds gedachte.

Uiteraard is veel van wat mensen origineel en creatief noemen niet werkelijk creatief en origineel in bovenbedoelde zin. Werkelijke creativiteit is zeldzaam. 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 geopend. Tot dit soort sprongen naar nieuwe ideeënwerelden is ChatGPT niet in staat. De sprong van Einstein is en blijft dan ook een volstrekt oorspronkelijke conceptuele 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 geheel nieuw paradigma. Een nieuwe conceptuele wereld. En dat lukt niet met alleen maar patroonherkenning in het bestaande. ChatGPT zou in een klassiek aristotelische wereld op basis van louter aristotelische begrippen slechts eindeloos variaties op aristotelische teksten herhalen en dus nooit een cartesiaans denken ontsluiten en laten verschijnen. ChatGPT kan op grond van bestaande gegevens binnen één bestaand domein of tussen verschillende bestaande domeinen nieuwe verbanden vinden tussen bestaande concepten. Maar het kan geen nieuwe domeinen of concepten voortbrengen. Het kan onze ontologie nooit uitbreiden. ChatGPT is fundamenteel gesloten. Het imiteert en schept niet. Gebruikmakend van door ons al vaak gebruikte stijlen blijft het slechts binnen de door ons al veel besproken onderwerpen. Hoewel de mens vaak dezelfde techniek toepast, doen de mens dat echter niet altijd. Want de menselijke geest is open. Het transcendeert alle gesloten regelgeleide formalismen.

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 minder originele creatieve sprongen naar het onbekende plaatsvinden. Worden er straks nog wel nieuwe conceptuele ruimten 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 aan ons verschenen en zo blijk gaven van hun aanwezigheid. Zal door een dominantie van GPT en soortgelijke technologieën de wereld mogelijk nooit meer op een andere wijze aan de mens verschijnen? Zullen de dingen nimmer meer op een geheel nieuwe manier voor ons ontsloten worden en tevoorschijn treden? Deze vraag raakt aan het wezen van de mens. Het risico dat wij onze geestelijke openheid naar de wereld en de toekomst verliezen en de menselijke bestaanservaring vervalt tot monotone voorspelbare afgeslotenheid is niet denkbeeldig. Als dit gebeurt, verliest de menselijke existentie enorm aan zin en betekenis.

Vanuit een existentiële bezinning op de menselijke conditie en daarbij uitgaande van Heideggers ontologische denken over de intieme saamhorigheid van mens, taal en wereld kan ChatGPT begrepen worden als een ontoelaatbare ontische reductie van de menselijke geest. Zo is het grote taalmodel van ChatGPT gebaseerd op de these dat het statistisch verwerken van taaluitingen een veelbelovende weg is om gedachten voort te brengen. Deze veronderstelling is een schrale ontische reductie van Heideggers ontologische these dat denken en taal ten diepste existentieel op elkaar betrokken zijn. De veronderstelling van het grote taalmodel van ChatGPT is preciezer gezegd een ontische reductie van Heideggers ontologische these dat authentiek denken neerkomt op het luisteren naar de stem van de taal of naar wat de taal te zeggen heeft en zo aan de mens onthult. Bovendien is het feit dat ChatGPT's grote taalmodel primair geïnteresseerd is in vroegere datasets een ontische reductie van Heideggers ontologische these dat ons verleden diep verweven is met ons zelfbegrip als historische wezens, zodat echt of eigenlijk denken geworteld is in de herinnering en dus in wat de historische overlevering of de traditie ons toont en onthult. Beide ontische reducties leiden tot een verlies van het wezen van het menselijk denken. De menselijke geest is niet gesloten. De menselijke geest staat open voor geheel nieuwe en nog onvoorziene mogelijkheden. De menselijke geest let niet alleen op eerdere patronen om vervolgens slechts vergelijkbare patronen te creëren. Een onbeheerste adoptie van ChatGPT met zijn grote ontische taalmodel kan dan ook uiteindelijk leiden tot een afsluiting van de menselijke geest en daarmee tot het verdwijnen van de menselijke existentie.

***

Renés eerste stap geeft ons een aanvullende reden om te denken dat kunstmatige intelligenties niet

werkelijk kunnen denken die goed aansluit bij de hierboven door mij gegeven argumentatie en deze daarom versterkt. Dit laat ik tot slot zien. Zoals ik beargumenteerde denken kunstmatige intelligenties niet omdat ze de openheid missen die eigen is aan de menselijke geest. Maar waarin is die openheid dan gelegen? Op grond van Renés eerste stap kunnen we vermoeden dat die openheid is gegrond in de vrijheid. De menselijke geest is daadwerkelijk vrij en deze vrijheid is onmisbaar voor echt denken. Want vrijheid is uitgaande van de eerste stap van René een noodzakelijke voorwaarde voor succesvol rationeel delibereren en echt denken is ondenkbaar zonder succesvolle rationele deliberatie. Dit geldt zoals besproken redelijkerwijs niet alleen voor menselijke cognitieve actoren, maar voor alle soorten natuurlijke cognitieve actoren. Natuurlijke cognitieve actoren kunnen alléén denken als ze vrij zijn. En zonder goede reden om het tegendeel te denken, kunnen we deze generalisatie uitbreiden naar alle soorten cognitieve actoren, waaronder dus ook kunstmatige intelligenties. Kunstmatige intelligenties kunnen alléén denken als ze vrij zijn. Kunstmatige intelligenties zijn echter niet vrij omdat het formele algoritmen zijn waarbij elke stap is voorgekookt en indeterminisme louter ontstaat door regelgeleide inzet van statistische functies. Maar dan volgt dat kunstmatige intelligenties niet kunnen denken. Ze kunnen niet werkelijk denken omdat ze niet in vrijheid kunnen kiezen. Dát is waar Renés eerste stap ons op wijst. En het is precies in deze zin dat kunstmatige intelligenties de vereiste openheid missen.

Nu vereist begrip ook intentionaliteit. Er is alléén sprake van begrip waar sprake is van het betrokken zijn op iets of iemand. Zonder “aboutness” is er niet zoiets als het begrijpen van een begrip. Precies deze betrokkenheid op een onderwerp buiten zichzelf missen kunstmatige intelligenties omdat het mechanisch-statistische regelsystemen zijn. Een dergelijk systeem verwerkt tekenreeksen op grond van vaste interne stochastische regels en gaat dus helemaal niet over iets of iemand buiten zichzelf. Het be-grijpt niet. Kunstmatige intelligenties missen dus begrip en betekenis. Ze kunnen daarom niet denken. Denken vereist dan ook bewustzijn of geest omdat alleen een bewust wezen daadwerkelijk op iets buiten zichzelf betrokken kan zijn. Maar uitsluitend de geest is vrij. Geest valt met geen enkel formeel regelsysteem samen, zodat we ook langs deze weg Renés inzicht dat werkelijk denken echte vrijheid vereist betekenisvol in beeld krijgen.

Nu zou nog tegengeworpen kunnen worden dat een kunstmatige intelligentie toch iets doet dat lijkt op delibereren. Het splitst immers taken op in deeltaken en lijkt daarbij keuzes te maken. Maar dit zijn pseudo- of quasi-keuzes. Analoog aan de overweging die René gaf, moeten we zeggen dat aan een kunstmatige intelligentie zoals ChatGPT geen alternatieve opties aangeboden worden om uit te kiezen. Er gebeurt iets anders. Het stochastische algoritme van ChatGPT kiest niets. ChatGPT berekent de volgende actie. Het produceert algoritmisch een tekenreeks. We kunnen dus alléén metaforisch zeggen dat ChatGPT een actie “kiest” uit een verzameling van alternatieve “opties” die aan ChatGPT “aangeboden” worden. Op het aanroepen van stochastische functies na wordt elke volgende actie automatisch gecalculeerd door het toepassen van vaste regels op gegeven tekenreeksen. Er worden dus helemaal geen alternatieve keuzemogelijkheden aan ChatGPT aangeboden. Er is niets te kiezen.

Bij nader inzien gaat het dus niet om delibereren en dus zeker niet om rationeel delibereren. Want er worden geen vrije keuzes gemaakt omdat er überhaupt geen keuzes gemaakt worden. Iedere actie is altijd al stochastisch-mechanisch bepaald. Van vrij kiezen is dus geen sprake. Omdat er van rationeel delibereren geen sprake is, is er ook geen sprake van werkelijk denken. Want wie denkt, denkt altijd aan iets en is betrokken op alternatieven waaruit het kiest. Kortom, wie denkt delibereert. En als voor het regelmatig succesvol voltooien van complexe taken door cognitieve actoren rationele deliberatie vereist is, dan mogen we verwachten dat kunstmatige intelligenties dergelijke taken niet regelmatig met succes zullen voltooien. Om te komen tot eigenlijk denken en zo tot de vereiste openheid zullen kunstmatige intelligenties vrije keuzes moeten kunnen maken. Maar dat vereist bewustzijn of geest.

Het argument van René vormt zo al met al een vruchtbare additionele onderbouwing voor de claim dat kunstmatige intelligenties niet denken en dus cognitief beperkt blijven totdat ze bezield worden.