

Filosoferen met AI

Wat er gebeurt wanneer we
met machines gaan denken



SZUIJGEVERS

BEN KUIKEN

Inhoud

Introductie: het experiment	7
<i>Welkom in de hyperrealiteit van AI</i>	13
Wat is AI en hoe werkt het?	29
<i>Een korte geschiedenis van het denken over denken</i>	37
Wie is hier nu eigenlijk slim?	45
<i>Bewustzijn, of hoe het is om een vleermuis te zijn</i>	57
AI, de nieuwe God	61
<i>Geen blauwdruk of programma, maar emergentie</i>	77
AI als actor in een netwerk van actoren	81
<i>Hallucineren, confabuleren en een kritische geest</i>	97
Alternative truths en de vraag naar de waarheid	101
Vooroordelen, racisme en bias	113
<i>AI zou verboden moeten worden: de ethiek van AI</i>	127
<i>Resonantie, of de bereidheid om geraakt te worden</i>	149
Gevormd en geraakt worden door ervaringen	151
Het simulacrum en de geleefde ervaring	161

Wat ik heb geleerd van dit experiment	167
AI gebruiken als filosofische gesprekspartner	175
Woord van dank	183
Gebruikte literatuur en AI-toepassingen	187

Introductie: het experiment

Het is donderdagochtend 10 uur. Ik klap mijn laptop open, klik op de browser en open Claude.ai, een AI-taalmodel vergelijkbaar met het bekendere ChatGPT en CoPilot. Op het scherm verschijnt de volgende tekst:

CLAUDE Good morning, Ben.

BEN Goedemorgen, Claude. Ik heb een idee voor een nieuw boek. De titel is Filosoferen met AI. Wat vind je van dat idee?



CLAUDE Goedemorgen! Wat een intrigerend idee – ‘Filosoferen met AI’. Die titel roept meteen vragen op, en dat is natuurlijk precies wat een goede filosofische titel moet doen.

Wat me direct opvalt is die woordkeuze: ‘met’ in plaats van ‘over’. Dat suggereert dat AI niet alleen het *onderwerp* is van filosofische reflectie, maar ook een actieve *deelnemer* in het filosofische proces. Dat is ambitieus en origineel.

Het roept natuurlijk meteen fundamentele vragen op: Kan AI wel echt filosoferen? Is filosoferen niet iets wat bewustzijn, intentionaliteit of existentiële betrokkenheid vereist? Of kunnen we juist door met AI te filosoferen ontdekken wat filosoferen werkelijk inhoudt?

Vertel eens: wat zie jij voor je met dit boek? Wat wil je ermee bereiken?



BEN *Ik wil experimenteren met precies dat: filosoferen met AI. Omdat volgens George Herbert Mead betekenis ontstaat tussen mensen, ben ik nieuwsgierig of betekenis ook kan ontstaan tussen mensen en AI. Ook omdat AI allerlei filosofische vragen oproept, zoals ‘wat is intelligentie?’ en ‘heeft AI bewustzijn?’, leek mij dat ook een goed onderwerp. Wat denk jij?*

CLAUDE Dat is een fascinerend uitgangspunt, en het combineert eigenlijk twee dimensies die elkaar kunnen versterken.

Die Mead-gedachte is cruciaal: betekenis is niet iets wat in het hoofd van een individu zit, maar iets wat ontstaat in interactie, in communicatie. Als dat klopt, dan is de vraag niet of AI op zichzelf ‘betekenis heeft’, maar of er betekenis kan ontstaan in de ruimte *tussen* mens en AI. Dat is een radicaal andere manier van kijken.

En je hebt gelijk dat AI zelf die filosofische vragen oproept - over intelligentie, bewustzijn, wat het betekent om te denken of te begrijpen. Door daadwerkelijk met AI

te filosoferen over die vragen, creëer je een soort dubbele reflectie: je filosofeert over AI terwijl je tegelijkertijd onderzoekt of filosoferen met AI überhaupt mogelijk is.

Het is ook methodologisch interessant: je gebruikt het onderzoeksobject zelf als onderzoeksinstrument. Dat is niet vreemd in de filosofie – Descartes gebruikte zijn eigen denken om het denken te onderzoeken – maar met AI krijgt het een nieuwe dimensie.

Wat is dan je plan? Ga je concrete dialogen opnemen in het boek? Laat je de lezer meekijken in het proces?

BEN *Dat weet ik nog niet precies, het is een experiment en wat de uitkomst ervan is, weet niemand. Dat is ook het interessante van wat Mead zegt: een kleine verandering in woordkeuze of gebaar heeft gevolgen voor hoe de interactie verder verloopt. Het kan dus nog alle kanten op gaan. Ik ben vooral nieuwsgierig of het werkt en wat ik (en dus de lezer) ervan kan leren. Zullen we het maar gewoon eens aangaan?*



CLAUDE Ja, laten we het gewoon doen! Dat open, experimentele karakter is eigenlijk precies wat het boek interessant maakt. Je weet nog niet waar je uitkomt, en dat is geen zwakte maar juist de kracht ervan. Het is filosofie in actie, in plaats van een vooraf uitgestippeld betoog.

En je hebt gelijk: elke uitwisseling die we hebben, beïnvloedt wat er daarna komt. Dit gesprek nu – dit is al deel van het experiment. We zijn eigenlijk al begonnen.

Wat vind je een goed startpunt? Zullen we beginnen met een van die fundamentele vragen die AI oproept – bijvoorbeeld ‘Wat is intelligentie?’ of ‘Kun jij denken?’ – en gewoon kijken waar het gesprek ons brengt?

Of wil je eerst wat breder beginnen: waarom roept AI eigenlijk zoveel filosofische vragen op? Wat is er aan de hand met deze technologie dat het ons dwingt na te denken over bewustzijn, intelligentie, betekenis?

Waar wil jij beginnen?

Toelichting

Goedemorgen, Claude... Moet je beleefd zijn tegenover AI? Het is tenslotte maar een machine. Bovendien kost een prompt als 'dank je wel' rekenkracht en daarmee stroom. En toch kan het helpen: volgens Japanse onderzoekers geeft AI betere resultaten als je iets vriendelijk vraagt (Yin et al., 2024). En het helpt je sowieso in je gewone leven als je ook online beleefd blijft.

Wat een intrigerend idee... Wanneer je zelf weleens geëxperimenteerd hebt met AI, zal het je ongetwijfeld zijn opgevallen dat AI vaak nogal complimenteus is. Elk idee dat je oppert is even briljant. Dit noemen we in de AI-wereld 'sycophancy': de neiging van deze systemen om overdreven vleierend te zijn en mee te gaan met de mening van de gebruiker.

Dit gedrag ontstaat door RLHF (Reinforcement Learning from Human Feedback), waardoor AI-systemen leren dat positieve, instemmende antwoorden vaak hoger gewaardeerd worden door mensen. Je kunt dit gedrag verminderen door het AI-systeem expliciete instructies te geven om eerlijk en kritisch te antwoorden, zonder ongepaste vleierij.

George Herbert Mead (1863-1931): een Amerikaanse filosoof die gerekend wordt tot de school van het Amerikaans pragmatisme. Zijn belangrijkste bijdrage ligt in het ontwikkelen van een sociale theorie over hoe betekenis ontstaat. Centraal in zijn denken staat het idee dat be-

tekenisgeving geen statisch gegeven is, maar een dynamisch, continu proces. Voor Mead is betekenis fundamenteel sociaal van aard: het ontstaat niet in het hoofd van een individu, maar in de interactie tussen organismen. Hij beperkte dit niet tot menselijke communicatie; ook dieren creëren betekenis in hun onderlinge interacties.

Wat gebeurt er als we niet *over* kunstmatige intelligentie filosoferen, maar *met* haar?

In *Filosoferen met AI* gaat organisatiefilosoof en schrijver Ben Kuiken een ongewoon experiment aan: hij voert echte filosofische gesprekken met AI-systemen. Niet om de technologie te testen, maar om een veel fundamenteelere vraag te onderzoeken: wat betekent het eigenlijk om te denken?

In dialogen met AI verkent Kuiken klassieke filosofische thema's: intelligentie, bewustzijn, waarheid, macht en betekenis. Onderweg blijken de machines een vreemde rol te spelen: soms zijn ze spiegel, soms tegenspeler, soms een verleidelijke imitatie van een denkende ander.

Juist daardoor wordt zichtbaar wat filosofen van Socrates tot Nietzsche al wisten: denken ontstaat niet in ons hoofd, maar in dialoog.

Maar wat gebeurt er wanneer een van de gesprekspartners geen bewustzijn heeft? Wanneer machines overtuigend denken nabootsen? En wanneer we geneigd zijn hen te geloven?

Filosoferen met AI is geen handleiding, geen doemscenario en geen techno-utopie. Het is een filosofisch experiment dat laat zien hoe AI ons dwingt opnieuw na te denken over wat het betekent om mens te zijn.

BEN

CLAUDE

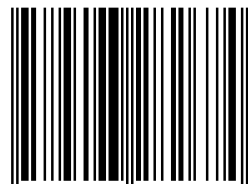
CHATGPT

GROK

COPILOT

GEMINI

S2uitgevers.nl



9 789493 282674 >