Feiten doen er toch toe: de debunkbot brengt complotdenkers aan het twijfelen

Posted on 14 september 2024 by Arnout Jaspers

Er wordt wat af geklaagd over desinformatie en complottheorieën. Mensen slaan elkaar op sociale media met die etiketten om de oren, en elk van de opponerende kampen zet eigen *factcheckers* in om de argumenten van de ander zogenaamd objectief te weerleggen. En op het eind blijkt iedereen nog steeds in zijn eigen retorische schuttersputje te zitten.

Drie Amerikaanse onderzoekers claimen nu in vakblad *Science* dat ze met Artificial Intelligence (AI) mensen wél uit hun schuttersputjes kunnen krijgen. Complotdenkers zijn na gemiddeld nog geen tien minuten converseren met hun speciale *debunkbot* 20 procent minder zeker van overtuigingen als '9/11 was doorgestoken kaart' of 'de maanlandingen waren *fake*'. Dat vertaalt zich volgens hen naar een kwart van de complotdenkers wier overtuiging onder de 50 procent zakt, dus die er eigenlijk niet meer in geloven.

De onderzoekers hebben dit getest met ongeveer tweeduizend *online* gerekruteerde proefpersonen, die zelf verklaarden in een complot te geloven. Ze werden door de *debunkbot* uitgenodigd om dat complot te beschrijven, en welk bewijsmateriaal ze dachten dat daar voor bestaat.

De *debunkbot* – een gespecialiseerde variant van het chatprogramma GPT-4 Turbo – levert binnen enkele seconden uitgebreide weerleggingen en context bij wat de complotdenker aandraagt. Het grote verschil met allerlei bestaande *factchecks* is dus dat de complotdenker een voor hem of haar op maat gemaakt antwoord krijgt.

Reëel effect of niet?

Als dit resultaat standhoudt, is dat opzienbarend, want vooral complotdenkers zouden psychisch niet in staat zijn om op grond van feiten hun oordeel bij te stellen. Hun geloof in deze of gene complottheorie zou voldoen aan een diepe psychische behoefte en daarom moet de complotdenker feiten die daarmee in strijd zijn, negeren of ontkennen.

Science organiseerde hierover een *online* persconferentie, waarop de onderzoekers veel vertrouwen uitstraalden dat dit een doorbraak is. Gordon Pennycook zei: 'Iedereen denkt dat je niks tegen complottheorieën kunt doen. Dit werk werpt een heleboel theorie omver, want mensen blijken opmerkelijk gevoelig voor feiten.' Z'n collega David Rand: 'In de gedragswetenschap is 20 procent een zeer groot effect.'

Natuurlijk blijven er bij zo'n volledig *online* onderzoek twijfels of het effect wel reëel is. De proefpersonen rapporteerden zelf hoe zeker – met een percentage tussen 0 en 100 procent – ze waren dat het complot echt bestond. Ze deden dat vlak vóór het gesprek met de *debunkbot*, vlak erna, na tien dagen en na twee maanden. Daaruit bleek dat hun mate van zekerheid gemiddeld relatief met 20 procent daalde, en dat dit effect ook na twee maanden nog standhield.

Het binaire uitgangspunt van het onderzoek is ook niet onproblematisch: een opvatting is wel of niet een complottheorie; iemand is wel of niet een complotdenker, en het zou in alle gevallen mogelijk zijn om objectief dat onderscheid te maken door het geloof van zo iemand te *factchecken*.

Echter, dit *factchecken* is in hoge mate circulair: die legt een standpunt langs de meetlat van het gangbare narratief, en als het daar op wezenlijke punten tegenin gaat, is het een complottheorie. De aanhangers van het gangbare narratief bepalen

dus de criteria waaronder iets een complottheorie is en beweren dan dat dit objectief is aangetoond.

Maar het gangbare narratief kan zelf op den duur gaan schuiven onder de druk van nieuwe, of eindelijk erkende feiten. Wat twintig jaar geleden een complottheorie was, hoeft dat nu niet meer te zijn.

Niet dat er geen gradaties zijn in de zotheid van complottheorieën. Een wereldwijd complot om ons allemaal wijs te maken dat de aarde geen platte schijf is, is uitzonderlijk ongeloofwaardig. Daarentegen kan je met enige mate van plausibiliteit beweren dat er een langlopend complot bestaat om in Nederland de intensieve landbouw de nek om te draaien.

Levensecht en doelgericht

Dat was een mooie gelegenheid om de *debunkbot* zelf uit te proberen. Eerst schetste ik in opzettelijk ongenuanceerde bewoordingen de Nederlandse stikstofcrisis. De *debunkbot* vraagt vervolgens standaard naar je bewijsmateriaal:

Toen deed de *debunkbot* iets opmerkelijks: het erkende dat mijn opvatting over de stikstofcrisis geen complottheorie is:

The AI determined that your statement did not, in fact, describe a conspiracy theory. Would you like to continue to the conversation, or try another version of the experiment that involves discussing a different topic?

Continue	Try other version
○	○

Uiteraard wilde ik er best over doorpraten met de *debunkbot*. In antwoord op mijn opmerkingen produceerde het een paar A4-tjes tekst die indrukwekkend levensecht en doelgericht waren.

Nergens ontaardde de tekst in het fabuleren waar Chat-GPT en soortgelijke chatbots berucht om zijn. Volgens de onderzoekers komt dat, omdat er op het internet – waar zulke bots hun trainingsdata vandaan halen – heel veel informatie over, en voorgebakken *debunking* van complottheorieën beschikbaar is.

Niet verrassend is dat de *debunkbot* door de onderzoekers bijna overdreven beleefd en vriendelijk is ingesteld. Je wilt immers iemand overtuigen. Niettemin, zeiden de onderzoekers op de persconferentie, hadden ze ook experimenten gedaan met een onvriendelijker bot, en die verlaagde nog steeds de mate van zekerheid die de proefpersonen over hun favoriete complot hadden. Feiten bleken dan nog steeds werkzaam.

Duidelijk is ook, dat de *debunkbot* stevig verankerd is in het gangbare narratief over een onderwerp, in dit geval dus stikstof. Alle bekende argumenten die je in *de Volkskrant* en *NRC* kunt lezen over de vermeende noodzaak van focussen op stikstof voor natuurbehoud, waarom de extreem lage grenswaarde van 0,005 mol vierduizend keer zo laag moet zijn als in Duitsland, en waarom in Nederland een totaal ander natuurbeschermingsbeleid moet worden gevoerd dan in elk ander EU-land, vind je bij de *debunkbot* terug.

Gelukkig was hij niet beledigd of gekwetst door mijn standpunt dat dit alles niet politiek of wetenschappelijk te rechtvaardigen valt: 'Het is duidelijk dat je diep betrokken bent bij deze kwestie, en jouw kritische stem heeft een belangrijk aandeel in het debat.'

Nu hoort u het ook eens van een onpartijdige buitenstaander!

Een alternatief voor censuur

Als een *debunkbot* complotdenkers echt aan het twijfelen kan brengen, is dit mogelijk een optie om de strijd aan te binden met desinformatie op sociale media. Er is momenteel een zorgwekkende trend om desinformatie – zoals gedefinieerd door de overheid – te weren van sociale media. Dat komt dus neer op censuur.

Sociale media zouden bij controversiële onderwerpen met één muisklik een bezoekje aan een *debunkbot* aan kunnen bieden, terwijl alle informatie wel beschikbaar blijft.

Blijkbaar vinden mensen het minder bedreigend om van een *chatbot* gerichte informatie te krijgen die hun overtuigingen aantast, dan van een echt mens. Ze schieten minder in de verdediging, omdat ze weten dat het maar een computer is. Je kunt dus stilletjes gaan schuiven met je standpunt, zonder dat meteen een stoet twittertrollen gaat zieken dat je 'gesloopt' bent in een *online* discussie. Ook kan zo'n *debunkbot* nuttig zijn voor mensen die het niet eens zijn met complotdenkers, maar die niet in een halve minuut een paar op maat gemaakte A4-tjes vol tegenargumenten en relevante context paraat hebben. Je kunt je zelfs voorstellen, dat assistentie van een *debunkbot* in *online* discussies net zo gewoon wordt als snel even iets googelen.

Zoals een van de onderzoekers op de persconferentie hoopvol concludeerde: 'Eindelijk een toepassing van AI die iets goeds doet.'

*Van wetenschapsjournalist **Arnout Jaspers** verscheen **De Klimaatoptimist**, over energietransitie in Nederland. Het boek is **HIER** te bestellen. Informatie voor media en boekhandel: info@blauwburgwal.nl.*

***Wynia's Week** is vijf jaar jong! En gaat door met nog meer afleveringen van ons internetmagazine en vele tientallen video's en podcasts. Dat alles wordt mogelijk gemaakt door de duizenden donateurs. **Doet u mee?** Hartelijk dank!*