

DE ACHILLESHIEL VAN LLM-AI

Niemand verbaast zich als minstens een kwart van een wetenschappelijk artikel uit foutenanalyse bestaat. En niemand zou een publicatie over waarnemingen waarin zo'n analyse van het experiment ontbreekt langer bekijken dan nodig is om op de prullenbakknop te klikken. Het is grotendeels door de rigoureuze foutenanalyse dat wetenschap niet 'ook maar een mening' is.

In het dagelijks leven is dat drastisch anders. Wie zegt "Die jas staat je goed", zou vreemd opkijken als de aangesprokene zou vragen om de hoeveelheid spreiding rond de aanduiding 'goed'. Dat komt door de taal. Die is sociaal zo bruikbaar dankzij een zekere mate van vaagheid. Maar daarom is het onmogelijk om een exacte vraag te formuleren, en een exact antwoord geven is evenmin mogelijk. Wij zijn dat gewend; sterker nog, taalvaagheid voelt goed. Daarom zijn de producten van AI zo aantrekkelijk. ChatGPT en aanverwante machinerie gebruiken grootschalige taalmodellen (Large Language Models, LLM) om correlaties in data, bijeengestolen door roofsoftware, aan de gebruiker te presenteren. In een vorm die voor iedere mens natuurlijk is: taal. Daarom zijn AI-producten onbetrouwbaar, en dat is iets anders dan onjuist. Ook het weerbericht voorspelt de temperatuur niet exact, maar je weet wel wat de verwachte afwijkingen zijn. Daardoor blijft die onjuistheid beheersbaar, en wordt de uitspraak betrouwbaar. Bij AI is dat niet het geval: je wordt bedonderd en je weet niet eens hoeveel.

Op een vraag aan AI krijg je nooit tweemaal hetzelfde antwoord. Dat wordt AI vaak verweten, maar dat slaat de plank

mis. Want ook in een wetenschappelijk experiment geeft een waarneming nooit exact dezelfde uitkomst als de vorige. Dat is nu juist een van de redenen dat foutenanalyse essentieel is. Zo'n analyse is bij AI principieel niet mogelijk, omdat het verschil tussen twee uitspraken niet meetbaar is, voornamelijk doordat taal symbolisch is. De afstand tussen Leiden en Haarlem is bekend, binnen een marge die ook bekend is. Maar wat is de afstand tussen twee complimenten? Twee dichtregels? Wat is de spreiding in de afstanden tussen twaalf partijprogramma's? 'Alle stukken aan stuurboord.' 'Alle hens aan dek.' Iedere matroos begrijpt het verschil tussen die twee, maar AI begrijpt niets, omdat er geen maat is voor de afstand ertussen. Grammaticaal is de afstand nul, maar als De Ruyter op AI had moeten varen, had ie de Vierdaagse Zeeslag geheid verloren. De vaagheid van taal, ingebakken in de LLM's, is de achilleshiel van AI. Hoewel ik denk dat deze fundamentele zwakte onoplosbaar blijft, is het theoretisch niet uitgesloten dat ooit een bruikbare kwantificering wordt gevonden voor taalkundige uitspraken. Dan kunnen we meten wat de afstand is tussen de ware uitspraak 'Vincent vertrouwt AI niet' en de mededeling 'Vincent is in 2026 jarig'. Ook dat laatste is waar, tenzij ik voor die tijd overlijdt... maar voorlopig zit ik nog binnen mijn foutmarge.



Vincent Icke
hoogleraar theoretische sterrenkunde aan
Universiteit Leiden en beeldend kunstenaar