Ingezonden brief in Vrij Nederland

Alexander Klöpping is op tv geweest, en wat volgt is vaste prik. Eerst appen vrienden en familieleden: ‘Zeg, is het echt zo gevaarlijk?’ ‘Gaat het echt zo hard?’, en de volgende dag heb je weer media aan de lijn die willen weten hoe het nu echt zit, en er met wisselend succes chocola van proberen te maken.

‘We hebben weer Klöpping-corvee’, zei Felienne Hermans, hoogleraar Computer Science Education aan de Vrije Universiteit Amsterdam, grappend in onze appgroep met AI-hoogleraren, techjournalisten en onderzoekers. Wij mogen weer komen opdraven om de schade zo goed en zo kwaad te herstellen.

Op magische wijze is Klöpping *de* techduider van Nederland geworden. Is er wat te vertellen, dan doet hij dat, voor een massapubliek. Logisch ergens, want hij heeft een vlotte babbel, en had tot voor kort altijd een positief verhaal over een toffe nieuwe app om je leven beter mee te maken. Bovendien heeft Klöpping, anders dan wetenschappers die zich in het debat mengen, niets te verliezen. Sterker nog: meer op tv komen is goed voor zijn *brand*.

Wetenschappers zijn, vaak terecht, huiveriger: voor je het weet leid je aan het zogeheten *Sagan Effect*: het idee dat wetenschappers, die net als astronoom Carl Sagan teveel aan publieke communicatie doen, minder serieus worden genomen. Of ze ondervinden er last van dat ze als te opiniërend worden gezien. Dat geldt ook voor deze brief. Er zijn wetenschappers die dit stuk alleen anoniem wilden steunen, uit angst voor gedoe op de campus.

‘Moet jij nou echt iedere week in de krant?’ werd een van ons laatst gevraagd. Nou nee, maar we voelen allemaal dat wij wel moeten. De hoeveelheid onzin die over het Nederlandse publiek wordt uitgestort, is namelijk niet te stelpen, en dat verziekt zowel het publieke debat als politiek beleid.

Grote Klöpping Show

Onlangs was het weer raak. Klöpping vertelde op 10 september bij de talkshow *Eva* over de kans dat AI de mensheid gaat uitroeien. Sommige wetenschappers, zei hij, hebben er zelfs een specifieke voorspelling bij. Aan bronvermelding doet hij niet, maar waarschijnlijk doelt hij op een medewerker van Anthropic die stelde dat de kans zo’n 10 procent is dat AI ons gaat uitroeien. 10 procent klinkt specifiek, maar hoezeer je ook zoekt: een onderbouwing is nooit te vinden.

De Grote Klöpping Show – waarin hij met een twinkeling in z’n ogen over het potentiële einde van de mensheid praat – zit vol met bekende antropomorfismes. AI wordt ‘vindingrijker’ en mag niet overleggen, maar ‘doet dat toch’ en ‘op eigen houtje’. Ze raken ‘in paniek’ bij de gedachte dat mensen ze in de smiezen kunnen krijgen. Maar goed, dat is toch gebeurd en op een bepaald punt trok OpenAI de stekker eruit, vertelt Klöpping voor een flitsende *visual*, voordat hij plaats mag nemen aan tafel om te zeggen dat ‘alles tegelijk gaat gebeuren en dat alles veel sneller zal gaan’. Doe maar niet te specifiek!

Uit een incident met taalmodellen volgt nog niet het sciencefictionverhaal van een AI die probeert te ontsnappen

Waarom wordt Nederland blootgesteld aan deze hijgerige hype? In één adem noemt hij het verlies van controle over de AI-agents, en het feit dat OpenAI het experiment stop kon zetten toen ze het ontdekten. En ja, een hack mede mogelijk gemaakt door taalmodellen is niet niks.

Maar uit zo'n incident volgt nog lang niet het sciencefictionverhaal van een AI die als zelfstandig handelend wezen probeert te ontsnappen, laat staan een kans van 10 procent op onze uitroeiing. De vraag of zulke scenario's niet wat overtrokken zijn, is dus terecht.

Zijn de modellen echt ontsnapt, of ontbrak het aan goede monitoring-*tools*? Een AI-agent, hoe geavanceerd ook, blijft altijd afhankelijk van gewone technische infrastructuur zoals het aanroepen van andere software, of toegang krijgen tot het internet via een netwerkverbinding. Dat zijn processen die wij als mensen kunnen controleren en waar nodig afsluiten. Internetverkeer kan niet in opdracht van een chatbot een onzichtbaarheidsmanteltje aantrekken.

Antropomorfische angstbeelden zijn juist makkelijk te visualiseren, zie alleen al die megaschermen achter Klöppings optredens, die gaan makkelijk viraal en nemen vervolgens alle zuurstof weg over de écht belangrijke vragen rondom digitalisering en AI. Waarom krijgt één tv-personality, zonder de juiste expertise, zoveel ruimte om een technisch, omstreden onderwerp te duiden, vrijwel zonder tegenspraak? Op welke basis krijgt hij die ruimte eigenlijk? Zijn prominente aanwezigheid in het publieke debat van Nederland ontneemt andere experts de ruimte om op basis van kennis van hun vakgebied duiding te geven. Om uit te leggen dat AI misschien wel krachtige codes kan genereren, maar geen menselijkheid heeft, geen intentie, en geen ziel.

Prullenbak omschoppen

Klöppings verhaal is eenzijdig. Telkens meldt hij dat AI slimmer is dan de mens, maar dat is een verhaal dat nuance verdient en aandacht vanuit allerhande vakgebieden. AI is als vraagstuk net zo veelzijdig als klimaatverandering. Het is een verhaal dat je kan belichten vanuit de meteorologie, maar ook vanuit historisch of politiek perspectief. Juist die breedte mist hier.

Jinek zou er ook voor kunnen kiezen om een filosoof of pedagoog te laten vertellen over wat creativiteit is, of intelligentie. Dat zijn geen simpele getallen die we kunnen meten, maar sociaal geconstrueerde begrippen. Zoals filosofe Mary Midgley uitlegt in haar boek *What is Philosophy for?*

In fact, the word 'intelligence' does not name a single measurable property, like 'temperature' or 'weight'. It is a general term like 'usefulness' or 'rarity'.

Een zin als 'AI wordt slimmer dan mensen' heeft, door deze bril, geen eenduidige betekenis.

Vraag een historicus om eens iets te vertellen over de chatbot ELIZA, waarin mensen in de jaren '60 op dezelfde wijze menselijkheid zagen, of over de Heider-Simmel-illusie, een experiment uit de jaren '40 waarbij mensen emoties toeschreven aan geometrische figuren.

Of vraag een psycholoog of psychiater iets te zeggen over hoe mentale processen in ons lichaam verankerd zijn. Want bedoelen de techbro's met 'slimmer dan een mens' ook iemands hand vasthouden, onbedaarlijk huilen, of in frustratie een prullenbak omschoppen? Soms is wanhoop zo groot dat dat de enige (dus de slimste) reactie is op verdriet of onrecht in de wereld.

Menselijkheid zit in een lijf, dát voelt spanning, liefde, honger. Paniek voelen, dat kan niet zonder lijf, want dat voel je in je hart, of in je longen of buik. Een AI-systeem kan dus niet 'in paniek' raken, want paniek is geen logisch gevolg van rekenkundige processen, maar iets dat je lichaam instinctief doet, vanzelf.

Als Klöpping weer op tv is, bijvoorbeeld.

Ondertekend,

Felienne Hermans, hoogleraar aan de VU en auteur van het boek *Vierkante ogen,* over het gebrek aan filosofische en historische perspectieven in de IT (verschijnt in oktober bij uitgeverij De Geus).

Antal van den Bosch, hoogleraar UU en AI-ontwikkelaar, auteur van het boek *Het geheim van de pratende machine* (Uitgeverij Prometheus), dat eveneens ik oktober uitkomt.

Iris van Rooij, hoogleraar in AI aan de RU, computationeel cognitiewetenschapper en projectleider Kritische AI-geletterdheid aan de RU.

Ilyaz Nasrullah, informaticus en digitaal strateeg, auteur van het deze week verschenen boek *Dit is AI: De valse belofte van kunstmatige intelligentie* (Uitgeverij Nieuwezijds).

Brenno de Winter, IT'er, informatiebeveiliger, ontwikkelaar en schrijver van *Fundamenten van Informatiebeveiliging* en *De validatiecrisis*.

Siri Beerends, cultuursocioloog, schrijver en onderzoeker bij medialab SETUP, promoveert binnenkort aan de UT in de Techniekfilosofie.

Emile van Bergen, maker van software voor beeldverwerking in medische apparatuur, autonome robots en andere *high risk*-systemen.

Piek Knijff, filosoof en data-ethicus, podcasthost van *Filosofie in actie*.

Laurens Vreekamp, journalist, panellid van Het Mediaforum (NPO Radio1) en mede-auteur van het boek *Niet onze AI: voorbij de beloften van Big Tech*.

Cynthia Liem, professor in Reponsible Digital Society aan Hogeschool Inholland, universitair hoofddocent Verantwoorde en betrouwbare AI aan de TU Delft.

Ajuna Soerjadi, filosoof en oprichter van het Expertisecentrum Data-Ethiek en schrijft aan het boek *De mythe van AI*.