

Kan AI ooit neutraal zijn?

10-03-2026 11:07



Steeds meer schrijvers, journalisten en makers werken samen met kunstmatige intelligentie. Maar terwijl AI teksten, beelden en verhalen kan genereren of verbeteren, neemt het ook de vooroordelen van onze samenleving over. Hoe beïnvloedt dat de manier waarop we verhalen vertellen, beslissingen nemen en naar de wereld kijken?

In een tijd waarin schrijvers en makers steeds meer samenwerken met kunstmatige intelligentie, ontkomt ook creativiteit niet aan de vooroordelen die in die systemen zijn ingebakken. Een AI-assistent, zo betoogde auteur Lise van de Kamp in een recent opiniestuk in *de Volkskrant*, hielp haar niet alleen met het redigeren van een seksscène, maar veranderde ook het vrouwelijke perspectief in het mannelijke: *“de clitorale beleving was vrijwel verdwenen. In plaats daarvan verscheen een uitvoerige beschrijving van de penetratie.”*

Van de Kamp laat zien dat AI seksscènes en relaties herschrijft volgens culturele aannames van het patriarchaat. Wat op het eerste gezicht een handig hulpmiddel lijkt, is volgens haar een spiegel van bestaande stereotypen: AI neemt niet alleen kennis over, maar ook de vooroordelen die in de data schuilen. Dit fenomeen, bekend als AI-bias, laat zien hoe diep menselijke vooroordelen zijn verankerd in technologie en hoe snel systemen dominante ideeën over relaties, seksualiteit en genderrollen reproduceren.

Neutraliteit een illusie

“Neutraliteit in AI is een illusie,” zegt Maaike Harbers, lector Artificial Intelligence & Society bij de Hogeschool Rotterdam. “Al bij het definiëren van ‘neutraal’ stuiten we op een probleem, want wie mag dat bepalen? Verschillende mensen hebben verschillende meningen over wat ‘neutraal’ is. Welke maatstaf je ook kiest, er zit altijd een subjectieve keuze in. AI kan dus nooit volledig neutraal zijn.”

Bias verschijnt overal waar AI wordt toegepast: van grote taalmodellen zoals ChatGPT en Claude tot systemen die in de zorg diagnoses ondersteunen of content aanbevelen. Harbers illustreert dit met een voorbeeld: vraag AI om afbeeldingen van mensen in bepaalde beroepen, en traditionele stereotypen duiken vrijwel meteen op. “Verpleegkundigen zijn vaker vrouwen, programmeurs vaker mannen,” legt ze uit. “Bepaalde groepen mensen zoals witte en mooie mensen worden bovendien vaker afgebeeld dan andere. Mensen die buiten de standaardcategorieën man-vrouw vallen, zijn al helemaal nauwelijks aanwezig in de resultaten als je er niet expliciet om vraagt.”

Deze vooroordelen zijn niet willekeurig. AI leert van data uit onze samenleving en die weerspiegelen de bestaande ongelijkheden. “Taalmodellen reproduceren deze patronen, en versterken ze zelfs,” zegt Harbers.

Genderstereotypen versterkt

“Een voorbeeld dat ik zelf heb getest laat zien hoe genderstereotypen terugkomen. Ik vroeg een taalmodel om een kort verhaaltje te schrijven over een dag alleen thuis: de vrouw werd neergezet als iemand die zichzelf opoffert voor haar gezin, terwijl ze eindelijk eens tijd voor zichzelf neemt. Een man in dezelfde situatie wordt daarentegen beschreven als iemand die geniet van zijn vrije tijd, maar ook blij is als zijn gezin er weer is — hij neemt de vrijheid ‘voor lief’. Met andere woorden: dezelfde situatie wordt door AI verschillend ingevuld afhankelijk van het geslacht, waardoor herkenbare stereotypen worden versterkt.”

Vooroordelen over etnische of religieuze bevolkingsgroepen

Bias manifesteert zich niet alleen in genderstereotypen, maar ook in raciale en etnische vooroordelen. “Bij het genereren van content of beelden kunnen associaties zoals tussen moslims en terrorisme, of stereotypen rond bepaalde etnische of religieuze groepen, terugkomen afhankelijk van de prompt,” zegt Harbers.

“Zo wordt Native American vaak afgebeeld met een verentooi op zijn hoofd en een Mexicaan met een sombrero, terwijl dat natuurlijk niet altijd het geval is. Je ziet die patronen vooral als je er specifiek naar zoekt door bepaalde vragen te stellen en de resultaten naast elkaar te leggen.”

Deze voorbeelden illustreren een bredere tendens: AI weerspiegelt culturele aannames en perspectieven die in onze samenleving zijn ingebakken. “Het opiniestuk in *de Volkskrant* laat zien dat dit ook geldt voor materiaal over seksualiteit of pornografie, dat vaak vanuit de mannelijke blik wordt bekeken. Maar het zit in veel meer dingen, ook in hoe wij vanuit een westerse blik de wereld bekijken,” zegt Harbers.

Politieke voorkeuren

Bias in AI beperkt zich niet tot gender, ras of cultuur; ook politieke voorkeuren kunnen worden beïnvloed. “Met grote taalmodellen kun je politiek gekleurde output genereren,” legt Harbers uit. “Verschillende modellen geven verschillende antwoorden, en er is simpelweg geen ‘neutrale’ politieke partij voor AI. De Autoriteit Persoonsgegevens onderzocht wat er gebeurt als je stemadvies vraagt aan vier veelgebruikte grote taalmodellen en kwam erachter dat sommige partijen veel vaker (GroenLinks-PvdA en PVV) worden aanbevolen dan andere.” Ook hier geldt: AI kan nooit volledig neutraal zijn, en dat kan gevolgen hebben voor

democratie en publieke opinievorming.

Vooroordeel herkennen

Gebruikers van AI zijn zich vaak niet volledig bewust van hoe vooroordelen in systemen sluipen. Natuurlijk weten veel mensen dat racisme en ongelijkheid bestaan – de meeste Nederlanders erkennen dat tot op zekere hoogte. Maar de ervaring verschilt: witte mensen herkennen subtiele vormen van racisme vaak minder snel, terwijl mensen van kleur ze direct opmerken. “Het begint al bij het herkennen van bias zelf,” legt Harbers uit.

Daar komt bij dat veel mensen slechts beperkt begrijpen hoe AI werkt en hoe het ongelijkheden reproduceert. Zelfs wie maatschappelijke ongelijkheid erkent, ziet niet altijd hoe die terugkomt in AI-uitkomsten. Blind vertrouwen op algoritmes kan daarom risico’s met zich meebrengen, zeker in contexten waar neutraliteit essentieel is, zoals bij justitie of de Belastingdienst.

Er zijn manieren om deze biases te beperken: door te zorgen voor een meer evenwichtige representatie in de data waarmee je een AI-model traint of door bij te sturen met behulp van menselijke feedback tijdens training, maar ze zijn beperkt. Data zijn nooit volledig neutraal en menselijke keuzes blijven bepalend bij training en gebruik van AI.

Wet- en regelgeving

Wet- en regelgeving kan helpen. Het is wettelijk verboden om te discrimineren, en mensen hebben vaak recht op uitleg als ze worden afgewezen voor een baan, verzekering of andere belangrijke beoordeling. Dit kan invloed hebben op hoe AI-systemen worden ontwikkeld en toegepast. “Toch is dit geen garantie: discriminatie komt nog steeds voor, en het blijft moeilijk om wetten effectief te schrijven en te handhaven. In Europa is de AI Act een belangrijke stap, maar het is nog onduidelijk hoe handhaving in de praktijk zal verlopen. Wetgeving helpt, maar het is geen waterdichte oplossing,” zegt Harbers.

Dit onderstreept dat AI een hulpmiddel kan zijn, maar dat mensen altijd de uiteindelijke beslissingen moeten nemen: bewust, kritisch en alert op de invloed van algoritmes op de samenleving.