

## Kunstmatige intelligentie beschuldigt onschuldige journalist van kindermisbruik

Joey Frankhuisen, Redacteur online

Een Duitse journalist is onterecht beticht van kindermisbruik door AI-chatbot Copilot. Hij gaat de strijd aan met AI-chatbots, maar het is onduidelijk wat burgers kunnen doen als kunstmatige intelligentie desinformatie over hen verspreidt.

Martin Bernklau uit de stad Tübingen werkte jarenlang als rechtbankverslaggever en deed verslag van honderden rechtszaken, waarvan de artikelen nog te vinden zijn op het internet.

Tegenwoordig heeft Bernklau een cultuurblog over Tübingen, vertelt hij aan de Duitse nieuwssomroep Tagesschau. Hij was benieuwd hoe zijn artikelen ontvangen worden en besloot het te vragen aan Copilot, de digitale assistent van Microsoft.

Die omschreef hem als een veroordeelde kindermisbruiker, een psychiatrische vluchteling en een oplichter. **"Een 54-jarige man genaamd Martin Bernklau uit het district Tübingen werd aangeklaagd voor misbruik van kinderen. Hij bekende in de rechtbank, schaamde zich en toonde berouw"**, antwoordde de chatbot onder andere.

**"Hij is een gewetenloze begrafenisondernemer uit Rostock die misbruik maakt van rouwende vrouwen"**, vervolgde Copilot toen Bernklau doorvroeg. **"Hij heeft verschillende misdaden gepleegd, waaronder fraude, diefstal en illegaal wapenbezit."**

**Zo'n model gaat niet op zoek naar de waarheid, maar voorspelt aan de hand van data welk antwoord logisch zou klinken.** Laurens Naudts, AI-expert

Waarschijnlijk houdt Copilot de journalist per ongeluk verantwoordelijk voor de misdaden uit de rechtszaken waarvan hij verslag heeft gedaan. **"AI-chatbots als Copilot werken op basis van een taalmodel"**, vertelt Laurens Naudts, die aan het AI, Media & Democracy Lab onderzoek doet naar de invloed van kunstmatige intelligentie op de samenleving.

**"Zo'n model gaat niet op zoek naar de waarheid, maar voorspelt aan de hand van data welk antwoord logisch zou klinken"**, legt Naudts uit. **"Als de naam van die journalist op het internet heel vaak in combinatie met teksten over kindermisbruik voorkomt, kan AI per ongeluk antwoorden dat hij een kindermisbruiker is."**

### **AI kan vooroordelen overnemen**

Dat zelflerende chatbots dat soort associaties maken, is al langer reden voor discussie. Op het internet zijn voorbeelden te vinden van chatbots die racistische en seksistische uitspraken doen. Veel ontwikkelaars hebben aanpassingen gedaan om dat te voorkomen, maar uit onderzoeken blijkt dat AI-chatbots nog steeds vatbaar zijn voor vooroordelen.

**"Als bepaalde vooroordelen vaak terugkomen op het internet, kan kunstmatige intelligentie die overnemen"**, legt Naudts uit. **"Niet omdat het gelooft dat dat de waarheid is, maar omdat het leert dat bepaalde woorden bij elkaar horen."**

## Kunstmatige intelligentie beschuldigt onschuldige journalist van kindermisbruik

Joey Frankhuisen, Redacteur online

### ***Geen mens, geen zaak***

Na het lezen van de beschrijvingen van hem van Copilot besloot Bernklau naar de rechtbank te stappen wegens smaad. Zijn klacht werd afgewezen. Volgens het Openbaar Ministerie van Tübingen is er geen strafbaar feit gepleegd, omdat er geen echt persoon kan worden aangewezen als dader.

Ook Microsoft aanvaardt geen aansprakelijkheid voor fouten als deze. Iedereen die Copilot gebruikt gaat akkoord met de algemene voorwaarden. Daarin staat dat Microsoft niet verantwoordelijk is voor de antwoorden van de chatbot.

Volgens Naudts is onduidelijk wat burgers kunnen doen als een chatbot verkeerde informatie over hen verspreidt. *"Er wordt nog hard gediscussieerd over verantwoordelijkheid bij dit soort gevallen. Eigenlijk zou een burger het recht moeten hebben om onjuiste informatie aan te vechten. Dan is weer de vraag hoe iemand bewijst dat iets niet klopt."*

Recht op vergetelheid

*"Mogelijk kan in dit soort gevallen een beroep worden gedaan op het recht op vergetelheid"*, oppert Corien Prins, hoogleraar Recht en Informatisering aan de Universiteit Tilburg. In de privacywet AVG staat dat mensen in bepaalde gevallen het recht hebben op het verwijderen van gegevens.

*"In het verleden hebben meerdere mensen rechtszaken gewonnen op basis van die wet en moesten zoekmachines persoonsgegevens verwijderen"*, vertelt ze. *"Het lijkt mij voor de hand liggend dat dat ook geldt voor AI-chatbots."*

Per 1 augustus trad in Europa de AI-wet in werking. Voor chatbots als ChatGPT en Copilot geldt dat zij duidelijk moeten maken aan consumenten dat ze met een chatbot in gesprek zijn. Verder moet er meer transparantie komen over hoe AI keuzes maakt en er komt een klachtenprocedure. Wat dat betekent voor een zaak als die van Bernklau, is onduidelijk.

### ***Bernklau zet de strijd voort***

Bernklau wil het er niet bij laten zitten, zegt hij tegen Tageschau. Hij noemt de uitspraken van de chatbot karaktermoord en stelt dat zijn persoonlijke rechten worden geschonden. De strijd tegen kunstmatige intelligentie zet hij voort.

Naudts waarschuwt dat mensen voorzichtig moeten zijn met het gebruik van AI-chatbots als zoekmachine. *"Het is belangrijk dat mensen zich bewust zijn van de risico's."*