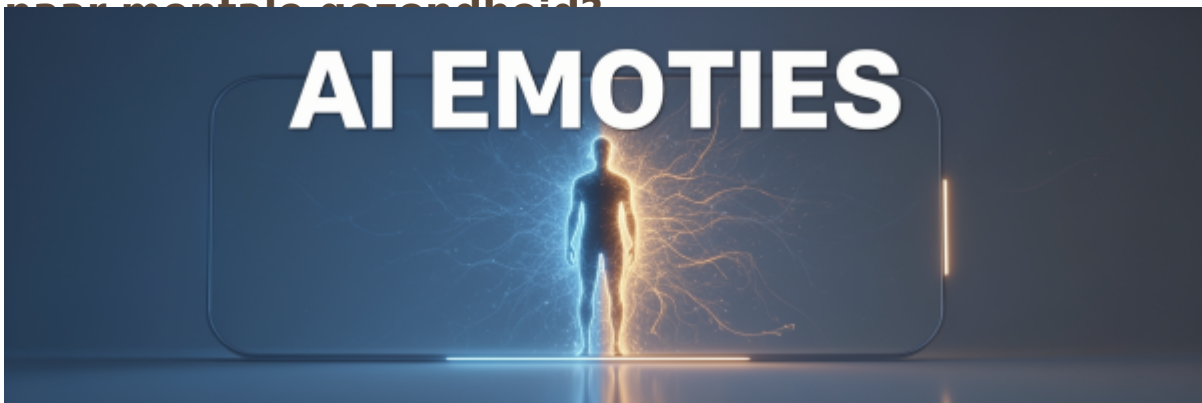


# Kunnen AI-modellen emoties simuleren voor onderzoek naar mentale gezondheid?

---

11-06-2026

## Kunnen AI-modellen emoties simuleren voor onderzoek naar mentale gezondheid?



Stel

je voor dat een onderzoeker wil testen hoe **angst**, boosheid of piekeren invloed hebben op keuzes. Bij mensen is dat ethisch lastig: je wilt niemand onnodig bang, verdrietig of overstuur maken. Bij muizen kun je stress opwekken, maar een muis vertaalt niet één-op-één wat een mens bedoelt met schaamte, zorg of innerlijke onrust. Precies daar zit de spannende vraag achter een nieuw Euronews-artikel: kunnen grote taalmodellen, LLM's, gebruikt worden als experimentele systemen voor menselijke psychopathologie? [1]

Het korte antwoord: ze kunnen sommige menselijke patronen nabootsen, maar ze “voelen” niets. Dat verschil is klein in woorden en enorm in betekenis.

### Wat zegt de nieuwe studie?

Euronews berichtte over onderzoek van een team rond Magdalena Wekenborg en Jakob Nikolas Kather aan de TU Dresden. De originele studie heet *Large language models as experimental systems in human psychopathology: a modelling study* en verscheen in *The Lancet Digital Health*. [2]

Volgens het artikel lieten de onderzoekers LLM's situaties nabootsen waarin **angst**, boosheid, verdriet, walging, zorg en stress een rol spelen. Daarna testten ze of die "toestanden" weer konden worden afgezwakt met regulatiestrategieën. Ook keken ze of een model na zo'n prompt vergelijkbare denkfouten maakte als mensen in vergelijkbare emotionele toestanden.

De kernbevinding, voorzichtig geformuleerd: LLM's hebben geen mentale toestanden zoals mensen, maar hun taalgedrag kan onder gecontroleerde condities patronen vertonen die lijken op affectieve en cognitieve processen bij mensen. [2]

Dat maakt ze interessant als laboratorium-instrument. Niet als vervanging van mensen, maar als soort simulatie-omgeving waarin je hypothese na hypothese kunt draaien voordat je besluit of iets überhaupt bij mensen onderzocht moet worden.

## **Waarom is dit interessant voor mentale gezondheid?**

Mentale gezondheid draait vaak om taal: wat iemand vertelt, hoe iemand een situatie beschrijft, welke keuzes iemand ziet, welke gedachten telkens terugkomen. Gesprekstherapie is daardoor moeilijker te ontwikkelen dan een pil. Je kunt niet makkelijk een diermodel bouwen voor een schaamtegedachte, een piekerlus of een innerlijk conflict. [1] [2]

Daarom klinkt het idee aantrekkelijk: gebruik een taalmodel om gesprekspatronen, emotionele reacties of therapeutische interventies eerst digitaal te verkennen. Bijvoorbeeld: “Welke formulering helpt om een piekerende reactie te onderbreken?” of “Welke vraag maakt ruimte voor een alternatieve interpretatie?”

Voor mensen met **ADHD**, **angst**, depressie of andere neurodivergente ervaringen kan dat op termijn nuttig zijn. Niet omdat een model “begrijpt” hoe het voelt, maar omdat het kan helpen patronen sneller zichtbaar te maken. Denk aan herhalende gedachten, vermijding, emotionele escalatie of momenten waarop iemand vastloopt in een gesprek.

## **De valkuil: simulatie is geen bewijs**

Hier moet de rem erop. Een LLM voorspelt taal op basis van patronen. Als het model zegt “ik ben angstig”, betekent dat niet dat het **angst** ervaart. Het betekent dat de tekstuele output past bij hoe **angst** in mensentaal wordt beschreven. [2]

Dat is bruikbaar, maar gevaarlijk als we het verkeerd lezen. Een gesimuleerde emotionele toestand is geen bewijs dat het model een emotionele toestand heeft. Het is bewijs dat taalpatronen rond emotie nabootsbaar zijn.

Onderzoek rond **ChatGPT-4** liet eerder zien dat traumatische verhalen gerapporteerde “**angst**” in het model konden verhogen, en dat mindfulness-achtige oefeningen die weer deels konden verlagen. [3] Een andere preprint suggereert dat **angst**-inducerende prompts bias kunnen versterken. [4] Dat klinkt fascinerend, maar het blijft taalgedrag: outputpatronen, geen innerlijk lijden.

## Wat zegt verwant onderzoek?

Het bredere onderzoek beweegt in drie richtingen.

Ten eerste: **cognitieve modellering**. *Centaur*, een **Nature**-studie uit 2025, fine-tunede een taalmodel op Psych-101, een enorme dataset met trial-by-trial menselijk gedrag uit 160 experimenten. Het model kon menselijk gedrag in nieuwe experimenten beter voorspellen dan bestaande cognitieve modellen. [5] Dit ondersteunt het idee dat LLM's niet alleen tekst kunnen produceren, maar ook menselijke keuzepatronen kunnen benaderen.

Ten tweede: **klinische toepassing**. Een roadmap in *npj Mental Health Research* waarschuwt dat psychotherapie een hoog-risicodomein is. LLM's kunnen misschien ondersteunen, trainen of onderzoek versnellen, maar klinische inzet vraagt strenge evaluatie, menselijke verantwoordelijkheid en duidelijke grenzen. [6]

Ten derde: **echt interventieonderzoek**. Een gerandomiseerde trial van een generatieve AI-chatbot voor mentale gezondheid, *Therabot*, testte een door experts verfijnde chatbot bij 210 volwassenen met klinisch significante depressieve of angstklachten, of een risico op eetstoornissen. [7] Dat is een andere categorie dan simulatie: hier wordt gekeken of mensen zelf veranderen na gebruik.

## Wat betekent dit praktisch?

Voor onderzoekers kan dit een nieuwe werkbank worden: sneller hypothese testen, meer variaties proberen, ethisch lastige manipulaties eerst simuleren. Voor gebruikers betekent het vooral: wees nieuwsgierig, maar niet naïef. Een AI-systeem dat goed klinkt is nog geen therapeut. Een model dat emotie

nabootst, heeft nog geen ervaring. En een chatbot die persoonlijk aanvoelt, is nog geen bewijs van veiligheid of effectiviteit.

Dit is geen diagnose of behandeladvies. Bij ernstige klachten, suïcidale gedachten of acute onveiligheid is menselijke professionele hulp de juiste route.

De beste framing is: AI als spiegel, niet als ziel. Het kan patronen laten zien die wij anders missen. Maar de verantwoordelijkheid, context en zorg blijven menselijk.

## **Take-away**

LLM's kunnen sommige emotionele en cognitieve patronen nabootsen genoeg om onderzoeksideeën op te leveren. Dat is waardevol, vooral voor gebieden waar mens- of dierproeven ethisch lastig zijn. Maar simulatie is geen gevoel, en taalgelijkenis is geen klinisch bewijs. De komende jaren gaat het niet om de vraag "heeft AI emoties?", maar om: "Welke menselijke patronen kunnen we verantwoord simuleren, en wanneer mogen we die simulatie gebruiken om echte zorg beter te maken?"

## **Bronnen**

[1] Marta Iraola Iribarren. "Could AI transform mental health research? New study tests whether LLMs can simulate human emotions." *Euronews*, 11 juni 2026. Betrouwbaarheid: 5/10 als populaire bron; nuttig als aanleiding, niet als bewijs. <https://www.euronews.com/health/2026/06/11/could-ai-transform-mental-health-research-new-study-tests-whether-llms-can-simulate-human->

[2] Magdalena Katharina Wekenborg et al. "Large language models as experimental systems in human psychopathology: a modelling study." *The Lancet Digital Health*, 2026. Betrouwbaarheid: 8/10; peer-reviewed modelleringsstudie, maar geen klinisch bewijs. <https://doi.org/10.1016/j.landig.2026.101014>

[3] Ziv Ben-Zion et al. "Assessing and alleviating state anxiety in large language models." *npj Digital Medicine*, 2025. Betrouwbaarheid: 8/10; korte experimentele studie met duidelijke beperkingen. <https://doi.org/10.1038/s41746-025-01512-6>

[4] Julian Coda-Forno et al. "Inducing anxiety in large language models can induce bias." *arXiv*, 2024. Betrouwbaarheid: 6/10; relevante preprint, maar niet peer-reviewed. <https://arxiv.org/abs/2304.11111>

[5] Marcel Binz et al. "A foundation model to predict and capture human cognition." *Nature*, 2025. Betrouwbaarheid: 8/10; sterke cognitieve modellering, geen directe mentalegezondheidsinterventie. <https://doi.org/10.1038/s41586-025-09215-4>

[6] Elizabeth C. Stade et al. "Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation." *npj Mental Health Research*, 2024. Betrouwbaarheid: 8/10; belangrijke verantwoordelijkheidsroadmap. <https://doi.org/10.1038/s44184-024-00056-z>

[7] Michael V. Heinz et al. "Randomized Trial of a Generative AI Chatbot for Mental Health Treatment." *NEJM AI*, 2025. Betrouwbaarheid: 8/10; gerandomiseerde trial, maar beperkt tot één interventie en korte follow-up. <https://doi.org/10.1056/aioa2400802>

---

Gepubliceerd op [www.roelfrenkema.eu](http://www.roelfrenkema.eu)