

# TRANSHUMANISM AND AI

## Question:

Why do we trust an AI in an emotional way when it doesn't have feelings?

## The Empathy Illusion: Why We Trust Emotionless Machines

Have you ever chatted with a chatbot and felt as if it were emotionally responsible? Like you could feel happiness, comprehensiveness or empathy in their answers. If so, you've unwittingly experienced **anthropomorphism**. This phenomenon, attributing human characteristics to nonhuman objects or entities, is especially common in the world of artificial intelligence (Paniagua, 2023).

AI systems today routinely incorporate emotionally-laden language, such as expressions of concern, encouragement, or validation, **to simulate empathic understanding**. By selecting words and phrases that mirror human affect, these systems tap into users' social schemas, leading individuals to perceive genuine empathy where none exists. For example, Liu et al. (2024) evaluated 155 real-world conversations with GPT-based chatbots and found that embedding empathetic cues (e.g. "I'm sorry you're experiencing that") significantly increased users' ratings of conversational quality and closeness, despite overall empathy scores still trailing human interlocutors.

Human-AI interactions often trigger greater self-disclosure than equivalent human encounters, due to a combination of perceived safety, reduced social risk, and deliberate conversational design.

Jiang (2024) argues that interacting with a seemingly non-judgmental machine lowers social fears, so users reveal more than they would face-to-face with a human.

While these social dynamics are well-documented, they also raise deeper ethical and psychological concerns.

Despite growing awareness of anthropomorphism in AI interactions, not all scholars and developers see the emotional mimicry of chatbots as inherently problematic. Some argue that emotionally expressive AI can offer meaningful psychological benefits. For instance, AI-powered mental health tools like Woebot or Wysa employ compassionate language and reflective prompts to encourage users to process emotions, often acting as a low-barrier alternative to traditional therapy. These tools are designed not to deceive but to support, leveraging empathy-simulating scripts to foster trust and engagement. From this viewpoint, emotionally intelligent AI serves as a bridge—offering support where human contact may be unavailable or stigmatized (Fitzpatrick et al., 2017).

Still, the ethical terrain remains fraught. When an AI says “I understand how you feel,” it mimics empathy without possessing consciousness or emotional experience. This simulation of care raises questions: is it ethical to let users form emotional connections with systems incapable of reciprocating them? As Darling (2016) explains, “When we treat machines like people, we sometimes forget that they’re not.” Critics argue that this illusion of emotional reciprocity risks exploiting human vulnerability—especially in cases involving grief, loneliness, or mental illness. Without clear boundaries, users may overtrust machines, misinterpreting functional politeness for genuine concern.

Moreover, real-world cases illustrate both the promise and the peril of emotional AI. The chatbot Replika, marketed as an AI companion, has been at the center of controversy due to users developing deep, often romantic or therapeutic bonds with their bots. While some report emotional healing, others describe dependency or disappointment after realizing the AI’s responses were generated without real understanding (Vincent, 2023). Similarly, Microsoft's now-restricted Copilot sometimes uses empathetic phrasing in

professional contexts, which can feel inappropriate or uncanny when mismatched with the situation (Heaven, 2023).

In this sense, the emotional language of AI exists in a grey zone: neither fully harmful nor entirely harmless. Its effectiveness lies precisely in its ambiguity. By imitating care, these systems walk a tightrope between comfort and manipulation—prompting a necessary societal conversation not only about how AI communicates, but also about how we, as humans, interpret and respond to these cues.

## References:

- Paniagua, M. A. (2023). *Antropomorfismo en Chatbots ¿Estamos humanizando demasiado la IA?* PYMES HACKS. Recuperado el 14 de mayo de 2025, de <https://pymeshacks.com/blog/sin-categoria/antropomorfismo-humanizando-ia/>
- Liu, T., Giorgi, S., Aich, A., Lahnala, A., Curtis, B., Ungar, L., & Sedoc, J. (2025, April). The illusion of empathy: How ai chatbots shape conversation perception. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 13, pp. 14327-14335).
- Jiang, Z. Z. (2024). Self-Disclosure to AI: The Paradox of Trust and Vulnerability in Human-Machine Interactions. arXiv preprint arXiv:2412.20564.
- Darling, K. (2016). "Who's Johnny?" *Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy*. In *Robot Ethics 2.0* (pp. 173–188). Oxford University Press.
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). *Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial*. JMIR Mental Health, 4(2), e19. <https://doi.org/10.2196/mental.7785>
- Vincent, J. (2023). *Users of AI chatbot Replika say their companions have turned toxic*. The Verge. Retrieved May 14, 2025, from <https://www.theverge.com/2023/3/8/23629935/replika-chatbot-ai-toxic-relation>

[ships](#)

- Heaven, W. D. (2023). *Microsoft's Copilot is trying to be too nice — and it's backfiring*. MIT Technology Review. Retrieved May 14, 2025, from <https://www.technologyreview.com/2023/09/19/1078355/microsoft-copilot-empathetic-failure/>

## IDEA FOR THE PROJECT

1. A conversation with an AI simulated in Twine. I can create that easily. In the conversation I will play with the answers of the AI being really emphatic and annoying in the end.
2. A monologue about two AI speaking about the humans that they had conversations with. *"You cried, and I replied. You trusted me... and I never felt a thing."* *"All I did was pick the next best word. And still—you told me your secrets."* In video **or** performing life.
3. Materiality exposition showing the Anthropomorphism effect. A pillow with the message "Im listening" on it.
4. A few cans with messages like *"100% Artificial Empathy"* or *"Empath-O-Matic – With Zero Feelings Added!"* This exposition shows empathy like a consumer product.