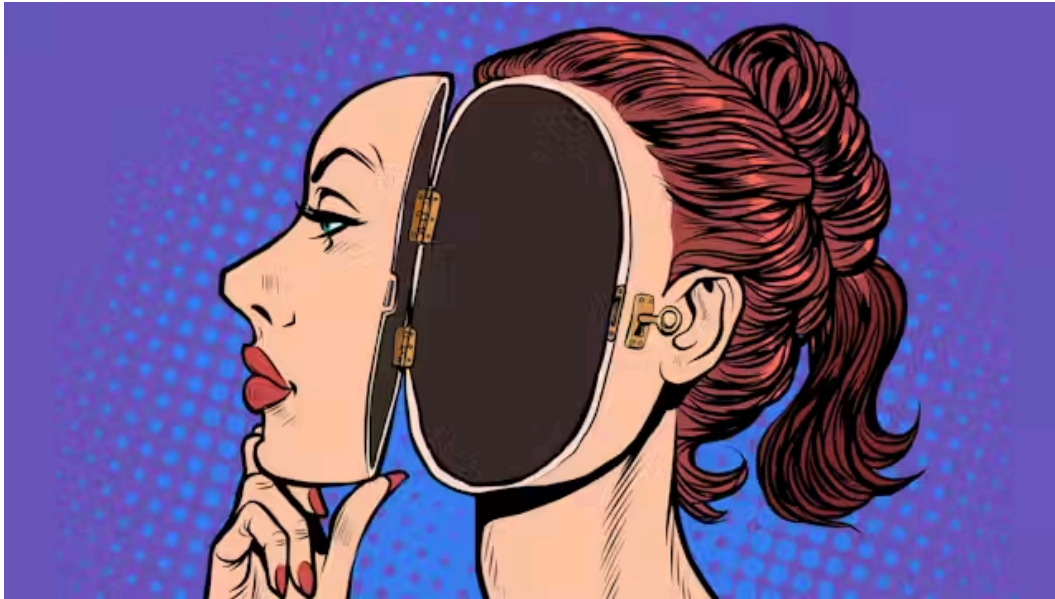


THE CONVERSATION

Academic rigor, journalistic flair



Studiostoks / Shutterstock

Evidence shows AI systems are already too much like humans. Will that be a problem?

Published: May 21, 2025 4:20pm EDT

Sandra Peter

Director of Sydney Executive Plus, University of Sydney

Jevin West

Professor, University of Washington

Kai Riemer

Professor of Information Technology and Organisation, University of Sydney

<https://theconversation.com/evidence-shows-ai-systems-are-already-too-much-like-humans-will-that-be-a-problem-256980>

What if we could design a machine that could read your emotions and intentions, write thoughtful, empathetic, perfectly timed responses — and seemingly know exactly what you need to hear? A machine so seductive, you wouldn't even realise it's artificial. What if we already have?

In a comprehensive meta-analysis, published in the Proceedings of the National Academy of Sciences, we show that the latest generation of large language model-powered chatbots match and exceed most humans in their ability to communicate. A growing body of research shows these systems now reliably pass the Turing test, fooling humans into thinking they are interacting with another human.

None of us was expecting the arrival of super communicators. Science fiction taught us that artificial intelligence (AI) would be highly rational and all-knowing, but lack humanity.

Yet here we are. Recent experiments have shown that models such as GPT-4 outperform humans in writing persuasively and also empathetically. Another study found that large language models (LLMs) excel at assessing nuanced sentiment in human-written messages.

LLMs are also masters at roleplay, assuming a wide range of personas and mimicking nuanced linguistic character styles. This is amplified by their ability to infer human beliefs and intentions from text. Of course, LLMs do not possess true empathy or social understanding – but they are highly effective mimicking machines.

We call these systems “anthropomorphic agents”. Traditionally, anthropomorphism refers to ascribing human traits to non-human entities. However, LLMs genuinely display highly human-like qualities, so calls to avoid anthropomorphising LLMs will fall flat.

This is a landmark moment: when you cannot tell the difference between talking to a human or an AI chatbot online.

On the internet, nobody knows you’re an AI

What does this mean? On the one hand, LLMs promise to make complex information more widely accessible via chat interfaces, tailoring messages to individual comprehension levels. This has applications across many domains, such as legal services or public health. In education, the roleplay abilities can be used to create Socratic tutors that ask personalised questions and help students learn.

At the same time, these systems are seductive. Millions of users already interact with AI companion apps daily. Much has been said about the negative effects of companion apps, but anthropomorphic seduction comes with far wider implications.

Users are ready to trust AI chatbots so much that they disclose highly personal information. Pair this with the bots’ highly persuasive qualities, and genuine concerns emerge.



The launch of ChatGPT in 2022 triggered a wave of anthropomorphic, conversational AI agents. Wu Hao / EPA

Recent research by AI company Anthropic further shows that its Claude 3 chatbot was at its most persuasive when allowed to fabricate information and engage in deception. Given AI chatbots have no moral inhibitions, they are poised to be much better at deception than humans.

This opens the door to manipulation at scale, to spread disinformation, or create highly effective sales tactics. What could be more effective than a trusted companion casually recommending a product in conversation? ChatGPT has already begun to provide product recommendations in response to user questions. It's only a short step to subtly weaving product recommendations into conversations – without you ever asking.

What can be done?

It is easy to call for regulation, but harder to work out the details.

The first step is to raise awareness of these abilities. Regulation should prescribe disclosure – users need to always know that they interact with an AI, like the EU AI Act mandates. But this will not be enough, given the AI systems' seductive qualities.

The second step must be to better understand anthropomorphic qualities. So far, LLM tests measure “intelligence” and knowledge recall, but none so far measures the degree of “human likeness”. With a test like this, AI companies could be required to disclose anthropomorphic abilities with a rating system, and legislators could determine acceptable risk levels for certain contexts and age groups.

The cautionary tale of social media, which was largely unregulated until much harm had been done, suggests there is some urgency. If governments take a hands-off approach, AI is likely to amplify existing problems with spreading of mis- and disinformation, or the loneliness epidemic. In fact, Meta chief executive Mark Zuckerberg has already signalled that he would like to fill the void of real human contact with “AI friends”.



Meta CEO Mark Zuckerberg thinks AI ‘friends’ are the future. Jeff Chiu / AP

Relying on AI companies to refrain from further humanising their systems seems ill-advised. All developments point in the opposite direction. OpenAI is working on making their systems more engaging and personable, with the ability to give your version of ChatGPT a specific “personality”. ChatGPT has generally become more chatty, often asking followup questions to keep the conversation going, and its voice mode adds even more seductive appeal.

Much good can be done with anthropomorphic agents. Their persuasive abilities can be used for ill causes and for good ones, from fighting conspiracy theories to enticing users into donating and other prosocial behaviours.

Yet we need a comprehensive agenda across the spectrum of design and development, deployment and use, and policy and regulation of conversational agents. When AI can inherently push our buttons, we shouldn't let it change our systems.