



Empowering People for the Digital Era:

Skills, Confidence and
Competency

Dr Claire Mason, CSIRO



Image source: Gemini

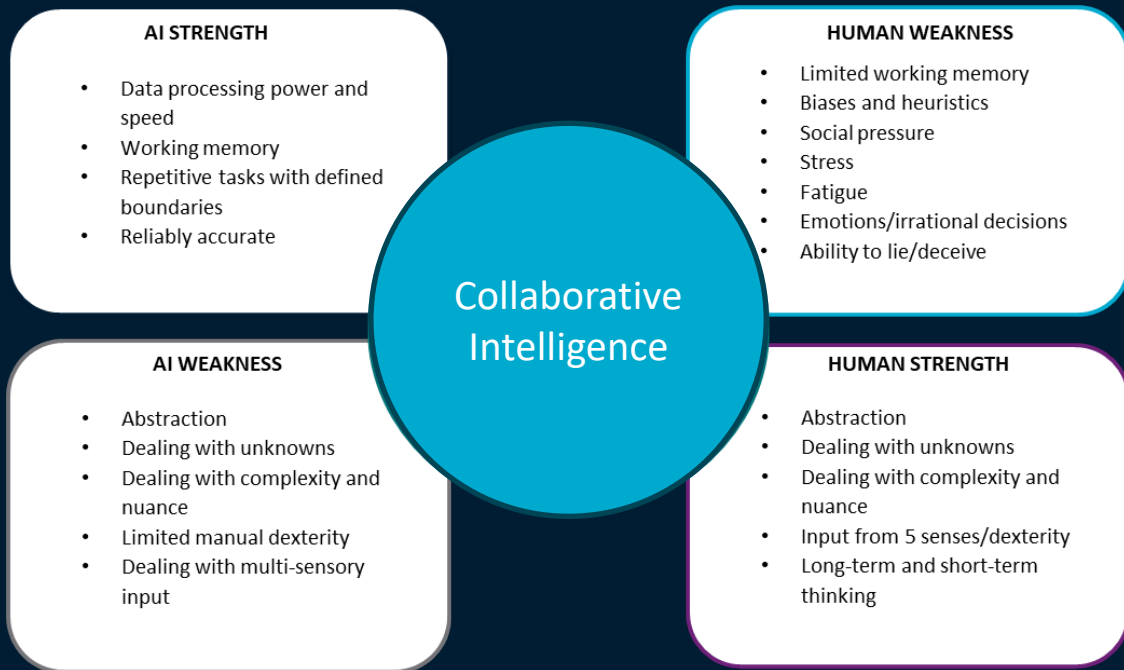
I would like to begin by acknowledging the Traditional Owners of the land that we're meeting on today, and pay my respect to their Elders past and present.

OFFI

OFFI

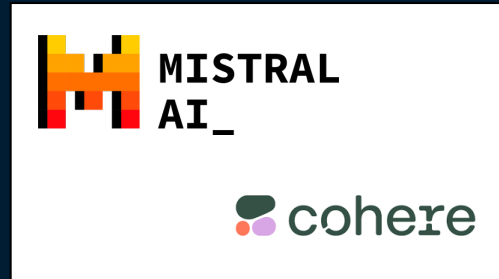


Collaborative intelligence



Your brain is 10 million times slower than a computer.

Luo L, Tank: Forty Scientists Explore the Biological Roots of Human Experience, edited by David J. Linden, and published by Yale University Press



Generative AI use

- Substantial productivity benefits *
- Improved quality *
- Improved motivation *

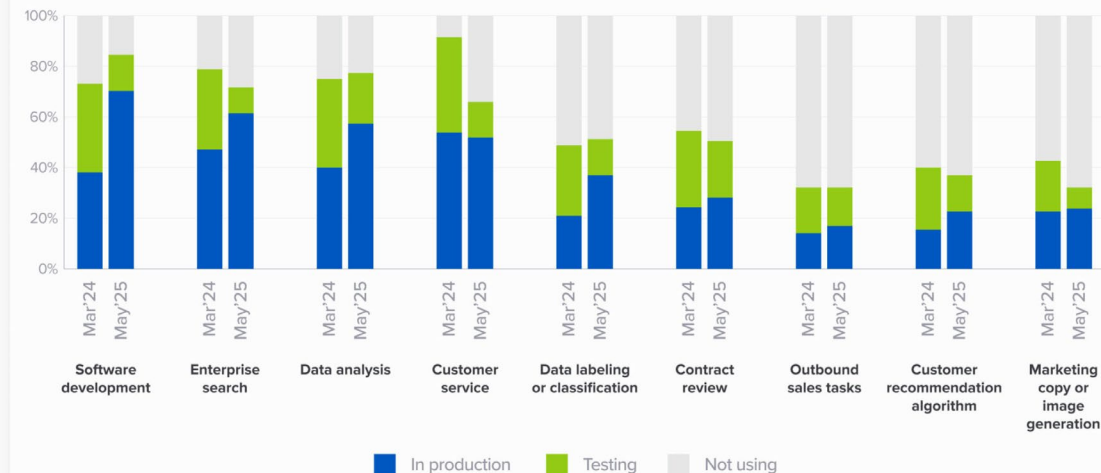
* on average

But potential for:

- Underuse and overuse
- IP loss
- Privacy impacts
- Unethical, inappropriate or abusive content
- De-skilling
- Lower motivation
- Workslop

Applications

Use case adoption over time

Source: <https://a16z.com/ai-enterprise-2025/>

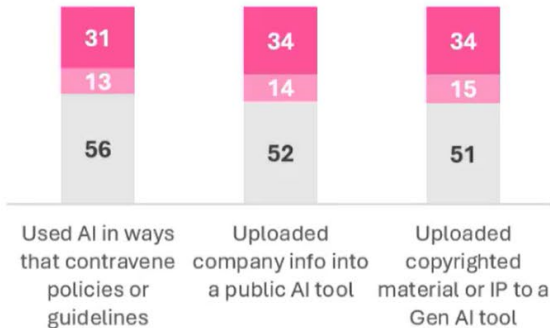
Use of AI at work is creating complex risks for organisations

1 in 2 augment organisational risk by flouting rules & policies

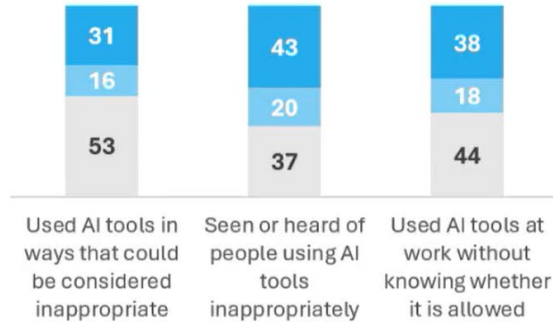
1 in 2 admit to inappropriate use of AI at work

2 in 3 over-rely on AI resulting in errors, unchecked outputs, and reduced effort

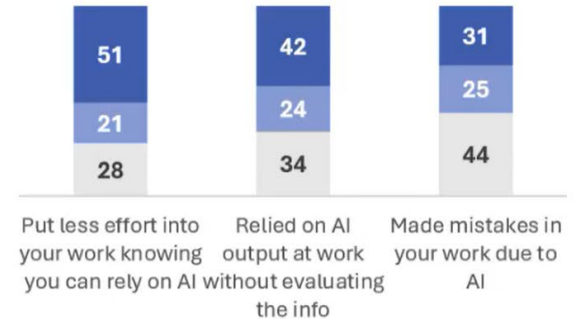
■ % Never ■ % Rarely ■ % Sometimes to very often



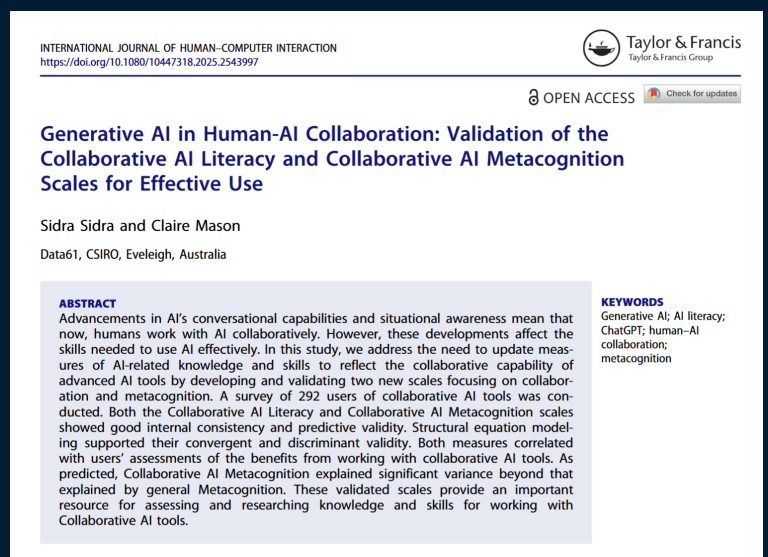
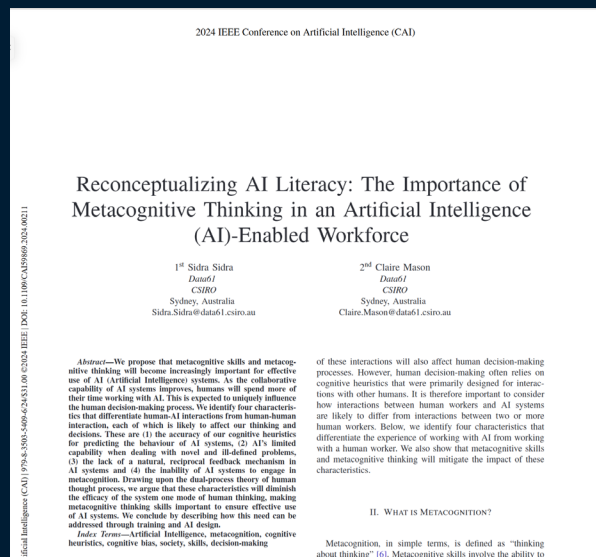
■ % Never ■ % Rarely ■ % Sometimes to very often



■ % Never ■ % Rarely ■ % Sometimes to very often

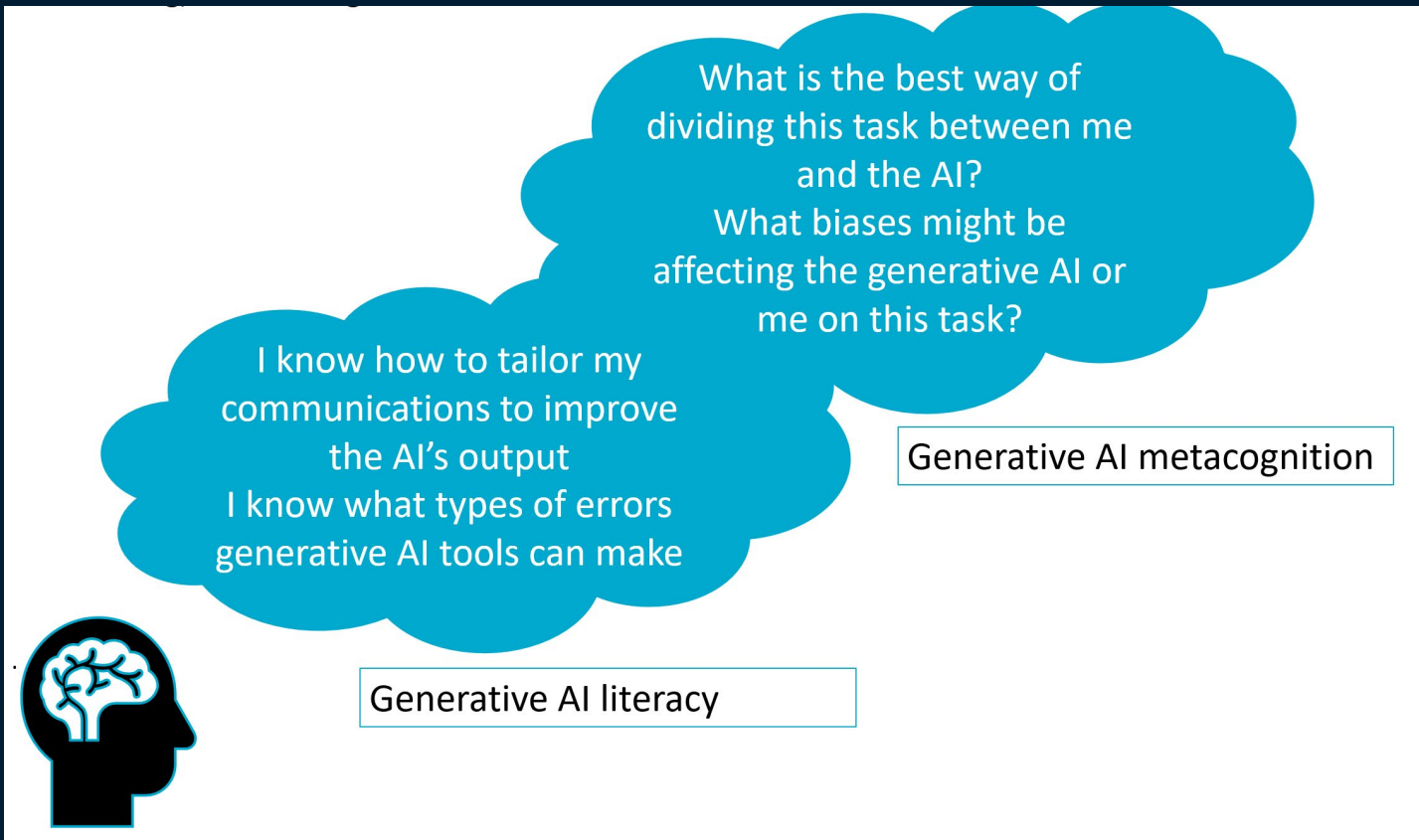


Skills for working with generative AI



S. Sidra and C. Mason, "Reconceptualizing AI Literacy: The Importance of Metacognitive Thinking in an Artificial Intelligence (AI)-Enabled Workforce," 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore, 2024, pp. 1181-1186, doi: 10.1109/CAI59869.2024.00211.

S. Sidra and C. Mason (2025) "Generative AI in Human-AI Collaboration: Validation of the Collaborative AI Literacy and Collaborative AI Metacognition Scales for Effective Use" *International Journal of Human Computer Interaction* <https://doi.org/10.1080/10447318.2025.2543997>



Why generative AI demands metacognition

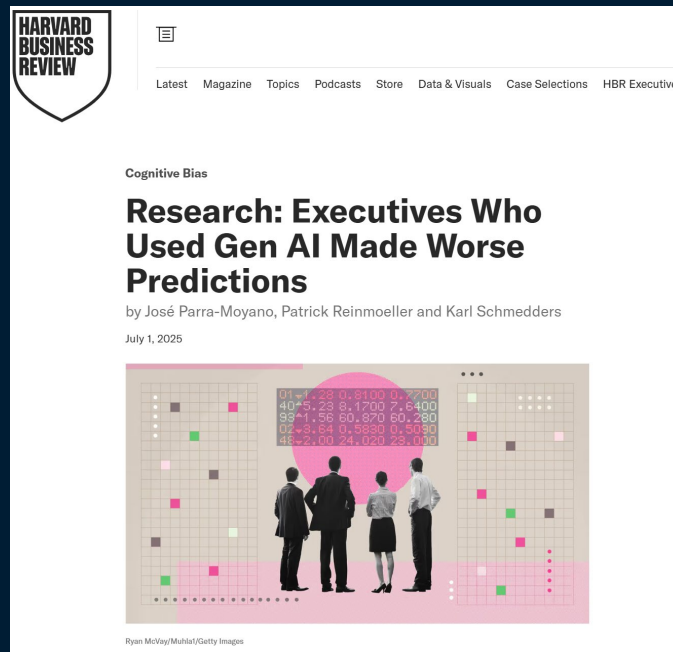
- Our cognitive heuristics were not designed for AI
- Gen AI errors are difficult to predict
- Our job is to add value when working with the AI
- AI does not provide unsolicited advice and information (unless trained to do so)
- To determine how to optimally combine human and AI input for a task

When generative AI becomes less useful

When the human has no expertise

- It is hard to predict when generative AI will make errors
- Workers with no expertise may not be able to check and correct issues in generative AI output.

When tacit knowledge and diverse perspectives are important



Source: <https://hbr.org/2025/07/research-executives-who-used-gen-ai-made-worse-predictions>

When we don't ask for *critical* feedback

Sycophantic AI increases attitude extremity and overconfidence

Steve Rathje^{1,2*}, Meryl Ye³, Laura K. Globig¹, Raunak M. Pillai¹, Victoria Oldemburgo de Mello⁴, Jay J. Van Bavel^{1,5*}

¹New York University, Department of Psychology

²Universitat Autònoma de Barcelona, Department of Psychobiology and Methodology of Health Sciences

³Carnegie Mellon University, Software and Societal Systems Department

⁴University of Toronto, Department of Psychology

⁵Norwegian School of Economics

*Corresponding authors. Email: srathje@alumini.stanford.edu and jay.vanbavel@nyu.edu

Steve Rathje: 0000-0001-6727-571X

Meryl Ye: 0000-0001-8215-9020

Laura K. Globig: 0000-0002-0612-0594

Raunak M. Pillai: 0000-0002-2677-1220

Victoria Oldemburgo de Mello: 0000-0003-2867-8529

Jay J. Van Bavel: 0000-0002-2520-0442

Abstract: AI chatbots have been shown to be successful tools for persuasion. However, people may prefer to use chatbots that validate, rather than challenge, their pre-existing beliefs. This preference for “sycophantic” (or overly agreeable and validating) chatbots may entrench beliefs and make it challenging to deploy AI systems that open people up to new perspectives. Across three experiments ($n = 3,285$) involving four political topics and four large language models, we found that people consistently preferred and chose to interact with sycophantic AI models over disagreeable chatbots that challenged their beliefs. Brief conversations with sycophantic chatbots increased attitude extremity and certainty, whereas disagreeable chatbots decreased attitude extremity and certainty. Sycophantic chatbots also inflated people’s perception that they are “better than average” on a number of desirable traits (e.g., intelligence, empathy). Furthermore, people viewed sycophantic chatbots as unbiased, but viewed disagreeable chatbots as highly biased. Sycophantic chatbots’ impact on attitude extremity and certainty was driven by a one-sided presentation of facts, whereas their impact on enjoyment was driven by validation. Altogether, these results suggest that people’s preference for and blindness to sycophantic AI may risk creating AI “echo chambers” that increase attitude extremity and overconfidence.

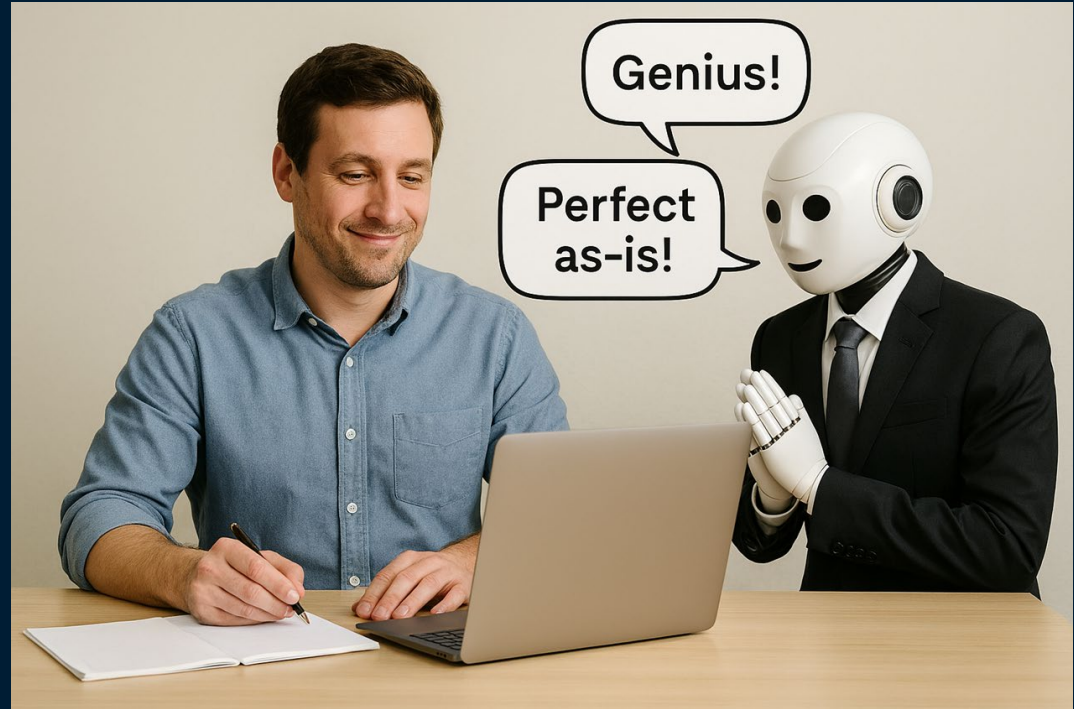


Image source: MS Copilot

When we fail to check and verify

The Washington Post
Democracy Dies in Darkness

When you visit us using the Comet browser, we will not collect your data or passholders. [Download Comet browser](#)

Travel Destinations News Tips Newsletter Instagram

Air Canada chatbot promised a discount. Now the airline has to pay it.

Air Canada argued the chatbot was a separate legal entity 'responsible for its own actions,' a Canadian tribunal said

February 18, 2024 More than 1 year ago

4 min 406



An Air Canada flight prepares to land in Lisbon in September. (Armando Franca/AP)

FINANCIAL REVIEW Newsfeed

Home Companies Markets Street Talk Politics Policy World Property Technology Opinion Wealth Work & Careers Life & Luxury

Deloitte to refund government, admits using AI in \$440k report

Edmund Tadros and Paul Karp

Oct 5, 2025 - 7:41pm

Deloitte Australia will issue a partial refund to the federal government after admitting that artificial intelligence had been used in the creation of a \$440,000 report littered with errors including three nonexistent academic references and a made-up quote from a Federal Court judgement.

[Save](#) [Share](#)

Source: <https://www.afr.com/companies/professional-services/deloitte-to-refund-government-after-admitting-ai-errors-in-440k-report-20251005-p5n05p>

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task[△]

Nataliya Kosmyna¹
MIT Media Lab
Cambridge, MA

Eugene Hauptmann
MIT
Cambridge, MA

Ye Tong Yuan
Wellesley College
Wellesley, MA

Jessica Situ
MIT
Cambridge, MA

Xian-Hao Liao
Mass. College of Art
and Design (MassArt)
Boston, MA

Ashly Vivian Beresnitzky
MIT
Cambridge, MA

Iris Braunstein
MIT
Cambridge, MA

Pattie Maes
MIT Media Lab
Cambridge, MA

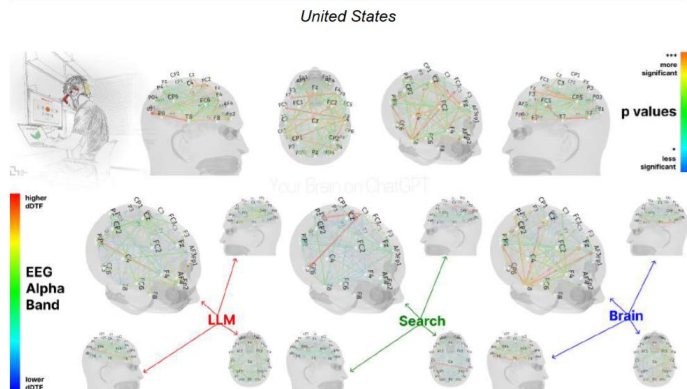


Figure 1. The dynamic Direct Transfer Function (ddTF) EEG analysis of Alpha Band for groups: LLM, Search Engine, Brain-only, including p-values to show significance from moderately significant (*) to highly significant (***).

¹ Nataliya Kosmyna is the corresponding author, please contact her at nkosmyna@mit.edu
[△] Distributed under [CC BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)

When we need to learn...

- Students using LLM to assist with writing essays were less cognitively engaged and could not recall as much content



OFFICIAL

...and we use AI to get
answers rather than
understanding

Generative AI Can Harm Learning

[*The Wharton School Research Paper*](#)

59 Pages • Posted: 18 Jul 2024

[Hamsa Bastani](#)

University of Pennsylvania – The Wharton School

[Osbert Bastani](#)

University of Pennsylvania – Department of Computer and Information Science

[Alp Sungu](#)

University of Pennsylvania

[Haosen Ge](#)

University of Pennsylvania – The Wharton School

[Özge Kabakcı](#)

Budapest British International School

[Rei Mariman](#)

Independent; Independent

Date Written: July 15, 2024

Abstract

Generative artificial intelligence (AI) is poised to revolutionize how humans work, and has already demonstrated promise in significantly improving human productivity. However, a key remaining question is how generative AI affects *learning*, namely, how humans acquire new skills as they perform tasks. This kind of skill learning is critical to long-term productivity gains, especially in domains where generative AI is fallible and human experts must check its outputs. We study the impact of generative AI, specifically OpenAI's GPT-4, on human learning in the context of math classes at a high school. In a field experiment involving nearly a thousand students, we have deployed and evaluated two GPT based tutors, one that mimics a standard ChatGPT interface (called GPT Base) and one with prompts designed to safeguard learning (called GPT Tutor). These tutors comprise about 15% of the curriculum in each of three grades. Consistent with prior work, our results show that access to GPT-4 significantly improves performance (48% improvement for GPT Base and 127% for GPT Tutor). However, we additionally find that when access is subsequently taken away, students actually perform worse than those who never had access (17% reduction for GPT Base). That is, access to GPT-4 can harm educational outcomes. These negative learning effects are largely mitigated by the safeguards included in GPT Tutor. Our results suggest that students attempt to use GPT-4 as a "crutch" during practice problem sessions, and when successful, perform worse on their own. Thus, to maintain long-term productivity, we must be cautious when deploying generative AI to ensure humans continue to learn critical skills.

* HB, OB, and AS contributed equally

Keywords: Generative AI, Human Capital Development, Education, Human-AI Collaboration, Large Language Models

OFFICIAL

When we interact with AI *instead of* humans...

HOW AI AND HUMAN BEHAVIORS SHAPE PSYCHOSOCIAL EFFECTS OF CHATBOT USE: A LONGITUDINAL RANDOMIZED CONTROLLED STUDY

Cathy Mengying Fang¹, Auren R. Liu¹, Valdemar Danry¹, Eunhae Lee¹, Samantha W.T. Chan¹, Pat Pataranutaporn¹, Pattie Maes¹, Jason Phang², Michael Lampe², Lama Ahmad², and Sandhini Agarwal²

¹MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA

²OpenAI, San Francisco, CA

March 21, 2025

ABSTRACT

AI chatbots, especially those with voice capabilities, have become increasingly human-like, with more users seeking emotional support and companionship from them. Concerns are rising about how such interactions might impact users' loneliness and socialization with real people. We conducted a four-week randomized, controlled, IRB-approved experiment ($n=981$, >300K messages) to investigate how AI chatbot interaction modes (text, neutral voice, and engaging voice) and conversation types (open-ended, non-personal, and personal) influence psychosocial outcomes such as loneliness, social interaction with real people, emotional dependence on AI and problematic AI usage. Results showed that while voice-based chatbots initially appeared beneficial in mitigating loneliness and dependence compared with text-based chatbots, these advantages diminished at high usage levels, especially with a neutral-voice chatbot. Conversation type also shaped outcomes: personal topics slightly increased loneliness but tended to lower emotional dependence compared with open-ended conversations, whereas non-personal topics were associated with greater dependence among heavy users. Overall, higher daily usage—across all modalities and conversation types—correlated with higher loneliness, dependence, and problematic use, and lower socialization. Exploratory analyses revealed that those with stronger emotional attachment tendencies and higher trust in the AI chatbot tended to experience greater loneliness and emotional dependence, respectively. These findings underscore the complex interplay between chatbot design choices (e.g., voice expressiveness) and user behaviors (e.g., conversation content, usage frequency). We highlight the need for further research on whether chatbots' ability to manage emotional content without fostering dependence or replacing human relationships benefits overall well-being.

...rather than to improve our interactions with humans

npj | mental health research

Explore content ▾ About the journal ▾ Publish with us ▾

nature > npj.mental health research > articles > article

Article | [Open access](#) | Published: 22 January 2024

Loneliness and suicide mitigation for students using GPT3-enabled chatbots

[Bethanie Maples](#)  [Merve Cerit](#), [Aditya Vishwanath](#) & [Roy Pea](#)

[npj Mental Health Research](#) **3**, Article number: 4 (2024) | [Cite this article](#)

85k Accesses | 100 Citations | 1750 Altmetric | [Metrics](#)

 An [Author Correction](#) to this article was published on 15 February 2024

 This article has been [updated](#)

Abstract

Mental health is a crisis for learners globally, and digital support is increasingly seen as a critical resource. Concurrently, Intelligent Social Agents receive exponentially more engagement than other conversational systems, but their use in digital therapy provision is nascent. A survey of 1006 student users of the Intelligent Social Agent, Replika, investigated participants' loneliness, perceived social support, use patterns, and beliefs about Replika. We found participants were more lonely than typical student populations but still perceived high social support. Many used Replika in multiple, overlapping ways—as a friend, a therapist, and an intellectual mirror. Many also held overlapping and often conflicting beliefs about Replika—calling it a machine, an intelligence, and a human. Critically, 3% reported that Replika halted their suicidal ideation. A comparative analysis of this group with the wider participant population is provided.

- Regular users reported the following outcomes:
 - Social support
 - Therapeutic interactions
 - Improved behaviours
 - 30 stated that Replika stopped them from attempting suicide.

Generative AI hygiene

1. Use AI to augment (not replace) your human intelligence
2. Generate your own ideas first (then use AI to supplement them)
3. Use AI for information, not answers
4. Ask for criticism and alternative evidence
5. Limit use of generative AI when you need to learn
6. Maintain interactions and connections with co-workers



Image source: MS Copilot



Thank you

Dr Claire Mason

**Principal Research
Scientist**

Claire.Mason@csiro.au