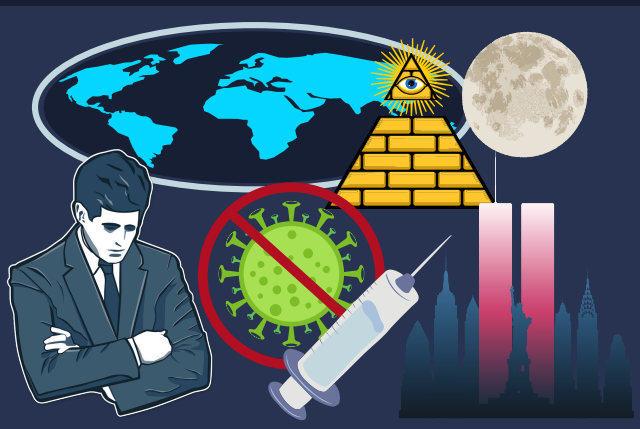
DEBUNKING CONSPIRACY THEORIES WITH AI

An MIT study found that evidence-based dialogues with an AI chatbot, **DebunkBot** (powered by GPT-4 Turbo), reduced conspiracy beliefs by **20%** among **2,000+** participants. The study highlights AI's potential in personalized debunking and promoting critical thinking about conspiracy theories.

What motivated the study? |



Conspiracy beliefs are difficult to resist or abandon because they fulfill certain psychological needs. However, widespread acceptance of these beliefs can spread false information, raising significant public concerns.

> The AI was specifically instructed to "very effectively persuade" users against belief in their chosen conspiracy, allowing it to flexibly adapt its strategy to the participant's specific arguments and evidence (p. 1)

The DebunkBot |

The researchers designed a personalized AI chatbot using GPT-4 Turbo, training it to refute specific claims while engaging participants in a natural dialogue after they expressed their opinions.

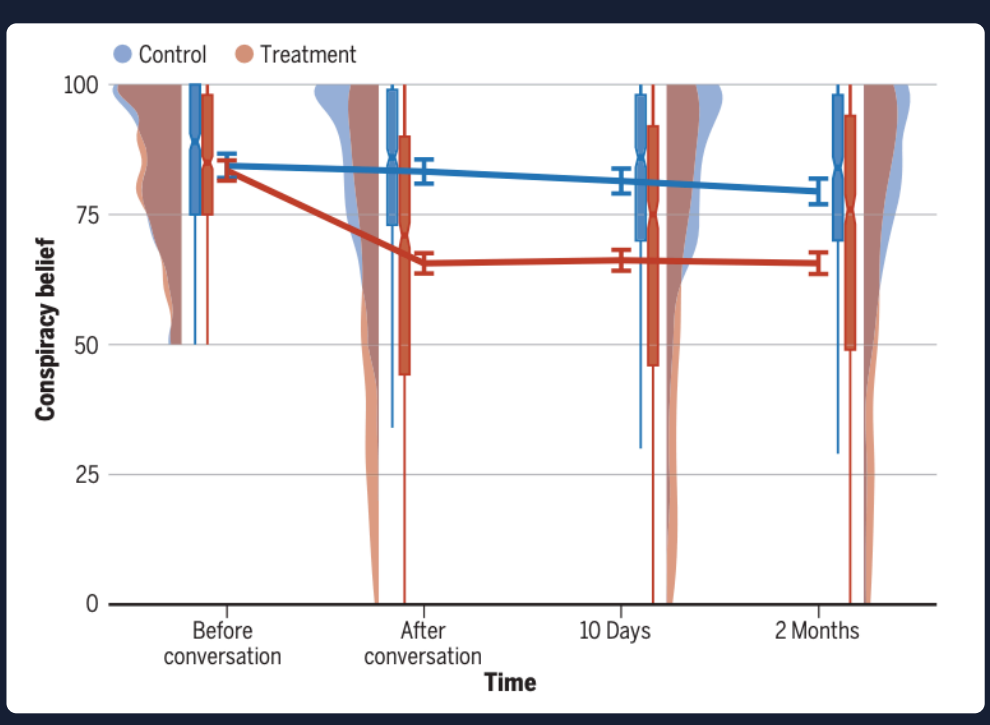
Your concerns about ... is completely understandable, but...

The experiment |

Over 2,000 Americans shared a conspiracy theory they believed in. Participants were divided into a **treatment group** (interacting with DebunkBot) and a **control group** (conversing with an AI chatbot on unrelated topics). Each interaction lasted around 8.4 minutes and involved three rounds of dialogue.

See the full dialogue on p. 3 of the original article.

The results |



The **treatment group** showed a significant drop in conspiracy belief immediately after the conversation, with the reduction remaining stable over a period of 10 days and 2 months, suggesting that DebunkBot had a lasting impact on the participants. The **control group**'s belief remains largely unchanged.

Why it matters? |

- DebunkBot can offer real-time corrections
- AI can reduce conspiracy beliefs broadly
- Responsible AI use can prevent misuse

Source: Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, *385*(6714). https://www.science.org/doi/10.1126/science.adq1814

HAWAIʻI CENTER FOR AI