

AI expert warns against telling your secrets to chatbots such as ChatGPT

Publication Date: 2023-12-26

Author: Jane Clinton

Section: Technology

Tags: Chatbots, Artificial intelligence (AI), Computing, ChatGPT, Royal Institution, BBC Four, news

Article URL: <https://www.theguardian.com/technology/2023/dec/26/ai-artificial-intelligence-secrets-chatbots-chatgpt-mike-wooldridge>



Confiding in ChatGPT about work gripes or political preferences could come back to bite users, according to an artificial intelligence expert. Mike Wooldridge, a professor of AI at Oxford University, says sharing private information or having heart-to-hearts with a chatbot would be “extremely unwise” as anything revealed helps train future versions. Users should also not expect a balanced response to their comments as the technology “tells you what you want to hear”, he adds. Wooldridge is exploring the subject of AI in this year’s Royal Institution Christmas lectures. He will look at the “big questions facing AI research and unravel the myths about how this ground-breaking technology really works”, according to the institution. How a machine can be taught to translate from one language to another and how chatbots work will be among the topics he will discuss. He will also address the question that looms around AI: can it ever be truly like humans? Wooldridge told the Daily Mail that while humans were programmed to look for consciousness in AI, it was a futile endeavour. AI, he said, “has no empathy. It has no sympathy”. “That’s absolutely not what the technology is doing and crucially, it’s never experienced anything,” he added. “The technology is basically designed to try to tell you what you want to hear – that’s literally all it’s doing.” He offered the sobering insight that “you should assume that anything you type into ChatGPT is just going to be fed directly into future versions of ChatGPT”. And if on reflection you decide you have revealed too much to ChatGPT, retractions are not really an option. According to Wooldridge, given how AI models work it is near-impossible to get your data back once it has gone into the system. A spokesperson for OpenAI, the organisation behind ChatGPT, said: “In April, we introduced the ability to turn off chat history. Conversations that are started when chat history is disabled won’t be used to train and improve our models.” Across the lecture series, Wooldridge will be joined by major figures from the AI world. The Royal Institution says he will also introduce “a range of robot friends, who will demonstrate what robots today can do – and what they can’t”. The Christmas lectures were started by Michael Faraday in 1825 at the Royal Institution in London with the aim of engaging and educating young people about science. They were first broadcast in 1936, making them the oldest science television series. Those who have given the lectures include the Nobel prize winners William and Lawrence Bragg, Sir David Attenborough, Carl Sagan and Dame Nancy Rothwell. The lectures will be broadcast on BBC Four and iPlayer on 26, 27 and 28 December at 8pm.