Hey! What's New? 2026-42

Would You Trust AI for Ethical Advice?

Most people would not prefer to take ethical advice from a computer, but a new Wharton study shows how attitudes change when users see how good AI guidance can be. An article on the *Knowledge at Wharton* webpage says that, in a blind study, even experienced professionals — including ethics professors and clergy — could not tell whether the AI-generated advice was more or less useful than the advice provided by Appiah. "In fact, random participants preferred the AI-generated advice 59.6% of the time. An additional study found that participants who were initially resistant to taking advice from the chatbot became more accepting as they were shown the high quality of the AI reasoning."

The team began by collecting 20 published questions and answers from the *The New York Times* column "The Ethicist." To eliminate differences in language style, they prompted GPT-4 with one example of the column author's writing so the AI verbiage would be similar. Then they assembled three groups of participants: a panel of ethics experts consisting of four pastors, a rabbi and 13 academics; a panel of 69 Wharton MBA students; and a panel of 100 people from Prolific, a platform that provides vetted participants for research.

According to the article, the groups were randomly shown the ethical questions and either the human response, the AI response, or both. "In either condition, the participants were blind to the response source. Then they were asked to rate which advice was more helpful. All three groups rated the advice high overall, and the groups exposed to both sources could not tell the difference between the AI and human responses.

The second study was designed to measure people's trust in AI. A different set of nearly 650 participants were put in three groups. The first group was given the ethical dilemma and asked if they would rather get advice from a human or computer. The second group got the ethical dilemma along with both the human and AI answers, with full disclosure on the source. Then they were asked which piece of advice was more useful. The third group was like the second, except they were blind to the source.

The results were telling, the article says. "In the first group, 72% said they would not prefer an AI response. That algorithmic aversion decreased to 53% in the full-disclosure group, and to 46% in the blind group. The preference for a human answer goes down a lot once they see the full source, because they see the quality of the AI. They no longer have this aversion."

According to Wharton professor Christian Terwiesch, such cases illustrate why AI must be developed and used responsibly. "I don't let a 6-year-old drive a Ferrari, and a 12-year-old has no business messing around with AI," he said. "This is the most powerful technology that humans have ever created, so people need guardrails to engage with it. It's not the fault of the machine. It's our responsibility as humans and as parents."

The scholars are continuing to experiment with what AI can do, including its ability to make accurate medical diagnoses. But for Terwiesch, who jokes that he caught a lot of flak over his, the ethics study is singular.

“My favorite part was when we hired these professors, and the rabbi, and my pastor friends,” he said. “It’s been one of the most fun academic studies I’ve done in my lifetime. But I do think that there is something big here: We all now have access to the highest quality ethical advice whenever we want to. We can ask the AI to provide us with different perspectives on a problem we face. How would a liberal deal with this dilemma? How about a conservative? What would the pope say? In this siloed world we live in, shouldn’t that give us the opportunity to overcome some of the divides we face and live a better life?”

Read more about this research at Would You Trust AI for Ethical Advice? - Knowledge at Wharton.