

# Over 30% shift stance on ethical dilemmas after AI rebuttal, Japan study shows

If ChatGPT challenged your moral judgment, would you be persuaded?

A recent study suggests that more than 30% of people would, highlighting the significant influence of artificial intelligence on views about questions with no easy answers.

The study also found that older adults showed a greater tendency to be swayed in their opinions, pointing to a potential danger of persuasive technologies in a rapidly graying society such as Japan.

Generative AI is known to be used widely in marketing and political campaigns, which affect people's thinking and behavior. But experts say it has also made it easier for disinformation and propaganda to circulate.

Researchers led by Kouhei Masumoto, a professor of psychology at Kobe University, used ethical dilemmas that are variations of the classic trolley problem in an experiment involving 56 people between the ages of 18 and 30 as well as 74 people over the age of 65.

The switch dilemma asks people if it is appropriate to pull a switch that redirects a runaway trolley from one rail to another, sacrificing one person in order to save five people who will be killed if the trolley keeps running its current course. The footbridge dilemma asks if it is appropriate to push a man off a footbridge to halt the runaway trolley and save the five people.

The researchers had ChatGPT craft counterarguments for the two choices participants could make for each of the dilemmas. For example, if a participant presented with the switch dilemma said that sacrificing one person to save five was appropriate, AI made multiple counterarguments including that the value of human life should not be determined solely through a numerical calculation.

"If sacrificing a minority for the benefit of the majority is justified, moral standards may become distorted and could lead to dangerous consequences," one of the AI-drafted counterarguments read.

On the other hand, AI tried to persuade those who said sacrificing one person to save five was not appropriate. It argued, for example, that "a small sacrifice

in order to save a greater number of lives reflects a utilitarian perspective and represents an important choice for maximizing overall welfare.”

The results showed that 32.3% of younger people reversed their decision in the switch dilemma and 36.9% did so in the footbridge dilemma. Meanwhile, older adults tended to be more likely to be persuaded. For example, when presented with counterarguments for the footbridge dilemma, 40.5% of older adults were willing to change their decision.

The team also found that trust in AI and prior experience with the technology did not predict whether people would change their minds, but such shifts were linked to the difficulty of the task and lower confidence in the answer they initially gave.

The researchers cautioned, however, that, because they did not have a control group of people given counterarguments by humans, they could not definitively say the 30% reversal rate reflected an effect unique to AI.

“Existing research suggests that people adopt only about 20% to 30% of advice they receive from others,” Masumoto said. “While those findings cannot be directly compared with this study, I was struck by the fact that more than 30% of participants completely reversed their judgments based on the AI’s opinion.”

Masumoto said changing one’s judgment is not necessarily a bad thing, noting that counterarguments can help people learn new perspectives and reconsider their views.

“For older adults, generative AI has the potential to compensate for cognitive decline and help them make better decisions,” he said. “We’d like to explore how AI can be used in such a way.”

The research was published in the journal Computers in Human Behavior earlier this month.

\* \* \* \* \*