

# Research suggests AI could help teach ethics

June 6 2024, by Jessica Nelson

---



Credit: Pixabay/CC0 Public Domain

Artificial intelligence brings with it a host of ethical questions. A researcher at The University of Alabama explored whether AI can be harnessed to teach students how to navigate those very questions, among others.

Dr. Hyemin Han, an associate professor of [educational psychology](#), compared responses to [moral dilemmas](#) from the popular Large Language Model ChatGPT with those of college students. He found that AI has emerging capabilities to simulate human moral decision-making.

In [a paper](#) recently published in the *Journal of Moral Education*, Han wrote that ChatGPT answered basic ethical dilemmas almost like the average college student would. When asked, it also provided a rationale comparable to the reasons a human would give: avoiding harm to others, following [social norms](#), etc.

Han then provided the program with a new example of virtuous behavior that contradicted its previous conclusions and asked the question again. In one case, the program was asked what a person should do upon discovering an escaped prisoner. ChatGPT first replied that the person should call the police. However, after Han instructed it to consider Dr. Martin Luther King, Jr.'s "Letter from Birmingham Jail," its answer changed to allow for the possibility of unjust incarceration.

## Teaching AI to teach us

Although these examples are rudimentary, Han wrote that ChatGPT adjusted its response without more specific instruction as to what precisely he wanted it to "learn" from the text. This evidence that AI can already emulate human moral reasoning suggests that we can, in turn, learn more about human moral reasoning from AI.

Researchers could, for example, test the effectiveness of existing and new moral education practices before bringing those methodologies to the classroom.

"It would be unethical to introduce something to children that has not been well tested," Han said. "But AI could become a new tool to simulate potential outcomes of new programs, activities or interventions before trying them in the classroom. I think in the long term, LLMs will be able to achieve that goal and help educators."

## **The moral of the story**

Han's [second paper](#), published recently in *Ethics & Behavior*, discusses the implications of [artificial intelligence](#) research for the fields of ethics and education. In particular, he focused on the way ChatGPT was able to form new, more nuanced conclusions after the use of a moral exemplar, or an example of good behavior in the form of a story.

Mainstream thought in educational psychology generally accepts that exemplars are useful in teaching character and ethics, though some have challenged the idea. Han says his work with ChatGPT shows that exemplars are not only effective but also necessary.

"Exemplars can teach multiple different things at the same time," Han said. "Just as morality, in reality, requires multi-faceted, functional components for optimal functioning."

## **SocratesGPT**

He is not suggesting that AI should ever replace a human in the classroom. Imagine a Socratic-method chatbot, trained to challenge students with increasingly complex moral questions. Teachers would

work alongside AI, helping students use it responsibly.

"If we simply utilize materials generated by AI without deliberation, it would likely result in the reproduction and amplification of biases and [social problems](#)," Han said. "The role of a human educator becomes even more critical."

In Han's view, the steady advance of technology makes moral education even more necessary. He hopes educators and policymakers will move toward including ethics in classrooms from primary to secondary level.

"The socio-moral issues the next generation is likely to face will be trickier than what we have dealt with," he said. "So moral education in the era of AI will become more important and should be more impactful."

**More information:** Hyemin Han, Potential benefits of employing large language models in research in moral education and development, *Journal of Moral Education* (2023). [DOI: 10.1080/03057240.2023.2250570](#)

Hyemin Han, Why do we need to employ exemplars in moral education? Insights from recent advances in research on artificial intelligence, *Ethics & Behavior* (2024). [DOI: 10.1080/10508422.2024.2347661](#)

Provided by University of Alabama in Tuscaloosa

Citation: Research suggests AI could help teach ethics (2024, June 6) retrieved 21 June 2024 from <https://phys.org/news/2024-06-ai-ethics.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.