596. Keep your enemies close: Embracing AI tools in AI ethics education

The generative AI boom of recent years has educators scrambling to find ways to deal with students using AI tools in assignments, as well as looking for ways to use the same tools productively and responsibly in their own work. Both the challenges and the opportunities are largely the same regardless of discipline, but there are also special cases. One such case is the teaching of AI ethics, where the challenges are, when viewed from another perspective, also opportunities. For the teacher, incorporating the use of AI tools in AI ethics courses can help illustrate the ethical implications of AI and the principles of responsible use. For the students, hands-on practice with AI helps understand and internalise the principles, so instead of banning the use of AI and treating it as a form of plagiarism, it may be more fruitful to actively encourage it and establish a set of ground rules. Failures to follow the rules can, at the teacher's discretion, be turned into teachable moments rather than used as grounds to deduct points or reject the student's work altogether.

In a pilot study carried out in 2024 in the Faculty of Information Technology and Electrical Engineering at the University of Oulu, generative AI was integrated into an AI ethics lecture course in three roles: as a tool for the teacher, as a tool for the student, and as a virtual tutor. The teacher, as part of an effort to redesign the learning assignments used on the course, used Microsoft Copilot for M365 to generate candidate topics for the assignments. The students were permitted to use AI tools of their choice to work on the assignments, provided that they declare the tools they used, document how they used them and reflect critically on the results. Documentation of the teacher's use of Copilot was made available to the students as a way of teaching by example. Finally, a chatbot designed to serve as a debating opponent was created using Azure AI Studio and deployed as a web application; as an optional bonus assignment, the students were instructed to discuss AI ethics with the bot and submit a chat log of the conversation.

The results of the pilot study were mixed but encouraging overall. When supplied with enough context in the prompts, Copilot was able to generate useful outputs, and the experience helped formulate instructions for meaningful and responsible use of generative AI for the students. However, there were several students who used AI without declaring it and when challenged, claimed not to have understood the AI policy or even been aware that such a policy was in force. This emphasises the importance of communicating the rules of AI use as unambiguously as possible and ensuring that the students accept them before submitting any coursework, possibly also involving the students in the specification of the rules to strengthen their commitment to following them. Nevertheless, the possibility of some students using AI without declaring it and without being detected remains an open issue.

The chatbot concept showed significant promise, requiring only a system prompt in natural language to specify the expected behaviour of the bot, which made it quick and easy to implement and deploy. Some two thirds of the students on the course completed the bonus

assignment, and the chat logs submitted by them demonstrated the ability of the bot to have meaningful conversations on a wide range of AI ethics topics. While primarily envisioned as a virtual person for the students to have debates with, the students independently discovered several other modes of interacting with the bot, and their feedback indicated a positive response to the bot as a way of supplementing human-human interactions on the course. Overall, while the results of the study are limited in their generalisability, they suggest that the appropriate attitude to adopt toward AI is one of skeptical but curious experimentation for co-creation of knowledge on the capabilities, limitations and ethical implications of the technology.

Authors of this abstract:

Lauri Tuovinen, University of Oulu