AI Education Matters: A Modular Approach to AI Ethics Education

Heidi Furey (Manhattan College; hfurey01@manhattan.edu)
Fred Martin (University of Massachusetts Lowell; fred_martin@uml.edu)
DOI: 10.1145/3299758.3299764

Introduction

In this column, we introduce our Model AI Assignment, A Module on Ethical Thinking about Autonomous Vehicles in an AI Course, and more broadly introduce a conversation on ethics education in AI education.

Why Ethics in an AI Course?

Recognition of the need for ethics education in the engineering-related disciplines goes back at least a hundred years (Layton Jr, 1986), but it's only been since the 1990s that expectations for ethics education have been adopted by accreditation bodies (Stephan, 1999).

The use of artificial intelligence and machine learning has become increasingly pervasive in our society, and this is accompanied by greatly expanding interest in the ethical considerations regarding their use. A crucial event was ProPublica's 2016 report investigating the use of AI tools by judges in criminal courts for determining the nature and lengths of sentences for convicts (Angwin, Larson, Mattu, & Kirchner, 2016). The article made a compelling case that the commercial software (which is in widespread use and employs secret algorithms) is racially biased.

International standards bodies are also tackling issues related to the ethical use of AI techniques, and specifically, the use of autonomous vehicles. The USA published a National Artificial Intelligence Research and Development (NAIRD) Strategic Plan, and the EU published a report, "Recommendations to EU Commission on Civil Law Rules on Robotics" (Martin & Makoundou, 2017).

Broadly, there is deep concern about the increasingly wide-reaching societal impact of AI approaches. It is urgent that our students are ready to think through the implications of their work, and make ethical choices.

Why A Modular Approach?

Here we are presenting a modular approach: a way of incorporating an introductory lesson about ethics into a one-semester AI course. Our module introduces students to the objectivist framework, and opens a specific conversation about the ethics of self-driving cars.

The modular approach is easily integrated into a single course. If the course is popular (or required), it reaches a large portion (or all) of the student population. Most importantly, students can connect the specific AI ideas they are learning to their ethical implications.

The limitations of the modular approach are mostly related to its short duration—there is only so much that can be accomplished with only one week of instruction.

A modular approach would be complementary with a whole-semester course on ethical thinking. Doing both would be more effective than only one or the other.

The Model AI Assignment

In our Module on Ethical Thinking about Autonomous Vehicles in an AI Course, we shared a set of resources for faculty use:

- Two days of lecture, in-class exercises, and discussion, introducing Utilitarianism and the Trolley Problem.
- A requirement that the course final project paper includes a discussion of the ethical implications of the project idea.
- A question on the final examination assessing students' understanding of the Trolley Problem.

These materials are elaborated in (Furey & Martin, 2018); here, we add further reflections and share more resources for introducing ethics to AI students.

Reflections on Ethics Teaching

One challenge to incorporating ethics into an AI course is helping students recognize the structure of ethical problems and their solutions. The field of ethics is both complex and far-ranging. Because of this, it can be difficult to meet this challenge within the space of an AI course without getting too far a field of the standard course material. With this concern in mind, we selected a topic that could serve as an example both of an ethical problem and ethical problem solving: the ethics of algorithm development for autonomous vehicles.

In the module, students were first introduced to the connection between ethical algorithms for autonomous vehicles and a classic ethical dilemma: the Trolley Problem. Afterword, students were guided through a worksheet driven group exercise designed to foster discussion and debate on the topic. This discussion helped prepare students to evaluate the potential benefits and shortcomings of an intuitive solution to ethical problems—Utilitarianism. Finally, students were offered examples of how one might construct a solution to these problems and challenged to continue thinking about the issue.

Thoughts on Implementation

The Trolley Problem presents a seeming intractable ethical dilemma — one in which every solution comes with an ethical cost. One benefit of exposing students to a somewhat eccentric philosophical example is that students become familiar with a key ethical problem-solving tool — the use of "thought experiment" — highly idealized hypothetical cases designed to test ethical theories and to isolate relevant moral variables. Another benefit is that students come to understand the complex nature of ethical dilemmas, and to recognize that solutions to such problems are rarely straightforward and may perhaps be equally complex. Too often, students who are unfamiliar with ethical problem solving resist thinking about ethical dilemmas because the answers appear "unknowable." Here, it is useful to draw parallels between difficult moral questions and difficult technical questions, reminding them that that complexity does not necessarily equal intractability.

It is useful to give them some general background in the field of ethics. For instance, It is important for students to recognize that there are different sorts of questions that one might ask regarding human behavior, only some of fall under the domain of ethical inquiry. Ethicists, as opposed to psychologists or sociologist, are interested in what people ought morally do to apart from what they in fact do or why they do it.

Other Resources

Burton et al.'s "Ethical Considerations in Artificial Intelligence Courses" (2017) provides an in-depth introduction to this conversation, including case studies and resource links. These authors also contribute the idea of using science fiction to engage students in thinking through ethical considerations (Burton, Goldsmith, & Mattei, 2015).

For a short video introduction to some ethics issues in AI, see Atlantic Magazine's Moral Code: The Ethics of AI.

For a variety of AI ethics resources including research, symposia, workshops, and reports, visit the AI Now Institute at New York University.

The AI Ethics Lab is a virtual organization that "brings together researchers and practitioners from various disciplines to detect and solve issues related to ethical design in AI."

Several articles from popular media address the challenges of introducing AI ethics into the classroom. For instance, see (Tugend, 2018) and (Holmes, 2018).

To date, more research has been done in engineering ethics pedagogy than in AI ethics pedagogy. For this reason, it can be useful to look for resources in engineering ethics. For a problem-solving approach to ethics instruction, see (Whitbeck, 1996). For a brief overview of some of the major ethical theories along with a discussion benefits and difficulties of teaching ethical theory, see (Bouville, 2008). For an article on teaching higher-order ethical concepts in engineering, see (Haws, 2004). For an example of an alternative to stand-alone ethics instruction in engineering, see (Riley, Davis, Jackson, & Maciukenas, 2009).

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*, *23*.

Bouville, M. (2008). On using ethical theories to teach engineering ethics. *Science and Engineering Ethics*, *14*(1), 111–120.

Burton, E., Goldsmith, J., Koenig, S., Kuipers, B., Mattei, N., & Walsh, T. (2017). Ethical considerations in artificial intelligence courses. *arXiv preprint arXiv:1701.07769*.

Burton, E., Goldsmith, J., & Mattei, N. (2015). Teaching AI ethics using science fiction. In *Aaai workshop: Ai and ethics.*

Furey, H., & Martin, F. (2018). Introducing ethical thinking about autonomous vehicles into an AI course. In *AAAI-2018.*

Haws, D. R. (2004). The importance of meta-ethics in engineering education. *Science and Engineering Ethics*, *10*(2), 204–210.

Holmes, W. (2018). The ethics of artificial intelligence in education. *University Business*.

Layton Jr, E. T. (1986). *The revolt of the engineers. social responsibility and the american engineering profession.* ERIC.

Martin, C. D., & Makoundou, T. T. (2017). Taking the high road ethics by design in AI. *ACM Inroads*, *8*(4), 35–37.

Riley, K., Davis, M., Jackson, A. C., & Maciukenas, J. (2009, March). "ethics in the details": Communicating engineering ethics via micro-insertion. *IEEE Transactions on Professional Communication*, *52*(1), 95-108. doi: 10.1109/TPC.2008.2012286

Stephan, K. D. (1999). A survey of ethics-related instruction in U.S. engineering programs. *Journal of Engineering Education*, *88*(4), 459-464.

Tugend, A. (2018). Colleges grapple with teaching the technology and ethics of A.I. *New York Times*.

Whitbeck, C. (1996). Ethics as design: Doing justice to moral problems. *Hastings Center Report*, *26*(3), 9–16.

Heidi Furey is an Assistant Professor of Philosophy and Director of the Ethics Center at at Manhattan College. Furey researches applied ethics and ethical pedagogy. Her current projects include work on the ethics of autonomous vehicles.

Fred Martin is Professor of Computer Science and Associate Dean for Student Success in the Kennedy College of Sciences at the University of Massachusetts Lowell. Martin develops novel technologies for CS and data science education, particularly for K–12.