Board 279: Ethics in Artificial Intelligence Education: Preparing Students to Become Responsible Consumers and Developers of AI

Dr. Helen Zhang, Boston College

Helen Zhang is a senior research associate working at the Lynch School of Education, Boston College. Her research interest includes STEM education, design thinking, and AI education.

Ms. Irene A. Lee, MIT STEP Lab

IRENE LEE is the PI of NSF ITEST Everyday AI and the NSF ITEST EAGER funded Developing AI LIteracy (DAILy) project.

Katherine Strong Moore, Massachusetts Institute of Technology

Kate Moore is a research scientist who studies how to teach middle and high school students about systems and ethics of artificial intelligence and machine learning. She earned her doctoral degree at Teachers College, Columbia University, where she studied cooperative learning and collaborative problem solving, and worked part-time as a professional development coach for STEM teachers in New York City public schools with the Center for the Professional Education of Teachers (CPET). Before entering the world of research and design, Kate served as a middle school science and special education teacher for 10 years. She has worked in public, independent, and charter schools in New York City NY, Newark NJ, and Pittsburgh PA.

Sheikh Ahmad Shah, Boston College

Sheikh Ahmad Shah is a 3rd year Ph.D. student in the Curriculum and Instruction Graduate Program at Boston College. His research primarily focuses on STEM education, scientific literacy, and AI literacy. He is currently working as a research assistant in the lab ”Innovation in Urban Science Education” led by Dr. Mike Barnett, Professor, Boston College. He also collaborates as a research assistant with Dr. Irene Lee’s team at MIT Media Lab on the ”Everyday AI” project.

Ethics in AI Education:
Preparing Students to become Responsible AI consumers and developers

Abstract

The rapid expansion of Artificial Intelligence (AI) necessitates a need for educating students to become knowledgeable of AI and aware of its interrelated technical, social, and human implications. The latter (ethics) is particularly important to K-12 students because they may have been interacting with AI through everyday technology without realizing it. They may be targeted by AI generated fake content on social media and may have been victims of algorithm bias in AI applications of facial recognition and predictive policing. To empower students to recognize ethics related issues of AI, this paper reports the design and implementation of a suite of ethics activities embedded in the Developing AI Literacy (DAILy) curriculum. These activities engage students in investigating bias of existing technologies, experimenting with ways to mitigate potential bias, and redesigning the YouTube recommendation system in order to understand different aspects of AI-related ethics issues. Our observations of implementing these lessons among adolescents and exit interviews show that students were highly engaged and became aware of potential harms and consequences of AI tools in everyday life after these ethics lessons.

Introduction

Artificial Intelligence (AI) is making an unprecedented impact on the industry and our society. While bringing much convenience and flexibility, the prevalence of AI in our life also leads to many unintended consequences. For instance, researchers have reported that bias in AI algorithms, which often emanate from unrepresentative or incomplete training data or the reliance on information that reflects historical inequalities, can result in flawed AI models. When these models are utilized to make inferences about people, such as facial recognition, predictive policing, and credit score assignment, they would lead to decisions which can have negative impacts on communities of color even without the programmer's intention to discriminate [1]–[3]. This has led to the ban on the use of such technologies in a few US cities. To empower young people to thrive in civic life in the era of AI, education must prepare them to understand the benefits and recognize potential harms of AI so that they can make informed decisions. However, this is not easy. Ethics is complex and requires critical thinking of perspectives of various stakeholders involved in the design of AI, which is often difficult for young adolescents as they tend to think in a more egocentric way [4].

Background

In the past decade, particularly in the past few years, researchers started to develop AI programs and curricula aimed at K-12 audiences, such as the MIT Responsible AI for Social Empowerment and Education (RAISE) initiative's collection of AI curricula and tools and

AI4All's Bytes of AI and full-length Open Learning curriculum for high school students. One emerging theme of many of these programs is that educators started recognizing the importance of teaching students about the ethical and societal dimensions of AI. For instance, the AI4K12 initiative [5], a collaborative effort between the Computer Science Teachers Association (CSTA) and Association for the Advancement of Artificial Intelligence (AAAI), developed a set of national guidelines for AI education for K-12 and included "Societal Impact" as one of the "Five Big Ideas of AI" that every K-12 student should know and be able to understand in AI. More recently, Long and Magerko [6] synthesized published work (2000-2019) into a conceptual framework of AI literacy and listed ethics as one of the core AI competencies, i.e., identifying and describing different perspectives on the key ethical issues surrounding AI (privacy, employment, misinformation, the singularity, ethical decision making, diversity, bias, transparency, and accountability).

Despite the increasing interest in teaching students about AI ethics, little is known how to teach or incorporate ethics related issues in AI curriculum. The traditional approach of teaching ethics as an isolated part in undergraduate computer science courses has failed to translate into experiences outside the classroom and left students unprepared for the current and future work in technology [7], [8]. Educators agreed that to prepare students to create ethical designs, ethics education needs to be embedded across the curriculum and engage students in practicing ethical decisions during the building of technologies. Yet there are still many debates about how to best accomplish the goals of ethics education, and the ways that different programs teach ethics are far from homogeneous in both content, pedagogy, and extensiveness [9].

Developing AI Literacy (DAILy): A curriculum featuring integrated AI ethics and technical learning

In our projects entitled "Developing AI Literacy" and "Everyday AI for Youth," we aim to develop and implement age-appropriate curricular materials and a teacher professional development program to develop AI literacy among middle school students. Our core curriculum, the DAILy curriculum, introduces AI concepts to youth through a socio-technical lens. The socio-technical lens is adopted by many engineering products to consider a design or product's potential impact on socio-technical systems, which span social, cognitive and information systems (i.e., hardware, software, personal and societal spaces) [10], [11]. For students, such a perspective can guide them to draw connections between their personal experiences with AI technologies, their communities, and potential impacts on the larger society of which they are a part.

Informed by research in engineering ethics education that much of the ethics instruction would run the risk of being only superficially effective if it does not address three categories of learning objectives: emotional engagement (want to make ethical decisions), intellectual

engagement (know how to make ethical decisions), and particular knowledge (be aware of the currently accepted guidelines for ethical practice) [12], [13], we curated a suite of ethics activities that expose students to various aspects of AI-related ethics issues and address learning of the three categories. Given that most middle school students have limited prior knowledge of ethics, these activities were designed following a carefully designed learning trajectory, which

1) stimulates students' ethical imagination through designing algorithms for making the "best" PB&J sandwiches and imagining the definitions of "best PB&J sandwich" by different stakeholders (e.g., parents, children, dentists). By creating these personas, students begin to understand that users' priorities can change the design of the algorithm;
2) helps students recognize ethical issues through investigating bias of existing technologies (e.g., Google Image search) and discussing whom the bias may impact;
3) helps students analyze key ethical concepts and principles that are applicable to the AI field (e.g., the Blueprint for an AI Bill of Rights) and encouraging them to take ethics seriously through case studies of how biased facial recognition technology harmed job applicants and misled police's judgments;
4) increases student sensitivity to ethical issues by hands-on experiments of training AI models using unbalanced datasets and playing games to understand how deepfakes and misinformation spread;
5) improves students' ethical judgment and willpower by engaging them in a culminating design project where they redesign the YouTube recommendation system. In these lessons, students critique the technology, identify its sources of bias (e.g., selective stakeholders in the design, datasets), and create a plan outlining how to improve the system.

Further, each ethics activity was designed following the lessons that teach related technical concepts to ensure that students possess adequate background technical knowledge in order to understand the ethics issues. For instance, Ethics lesson #3 was taught immediately after students learn the processes of supervised learning and experiment using Google's Teachable Machine to train AI models to detect faces. These ethics activities engaged students in reflecting on their personal and societal impact and brainstorming solutions to mitigate the harms, which contextualized the AI concepts and tools students learned, reinforced their learning of the technical aspects of AI, and highlighted the interrelatedness of technological tools with their human impact and societal implications.

Findings

We have implemented the DAILy curriculum in two online summer camps with a total of 58 middle school students. The campers met online for three hours every day for two weeks. All the activities were implemented synchronously via Zoom and all curricular materials were accessible through Google Classroom. Both camps were co-taught by a team of middle school

teachers who have learned the DAILy curriculum as learners and co-planned with experienced teachers on how to implement the curriculum. Each session typically started with the teacher introducing the unit's topic, followed by a whole-class activity, a small group or individual activity, and a discussion relating to ethical implications. Participants were randomly grouped into three groups of 7 or 8 individuals for small group discussions and hands-on activities.

A preliminary analysis of the observation data shows a high engagement of all students and an increase in their AI ethics awareness and knowledge. For instance, upon the completion of Ethics lesson #2 (investigating bias of existing technologies), one student described her takeaway as "*when we typically think of Google, we think it's objective, it's always right, but now I know it's not always represented in the right way.*" Another student reported that "*My takeaway is that AIs like Google and facial recognition, you can clearly see the bias in that. And even though we think that these things are very... very smart, but I never considered that they might have been really biased until going over these things [activity].*" This suggests that by exposing students to bias in everyday technology, students started recognizing the bias issue in technologies that they would normally consider as objective. After Ethics lesson #4 (experimenting with training AI models using unbalanced datasets), a student concluded that "*Larger dataset = more information for the AI to train with = better AI."* While his conclusion is not entirely accurate (an ideal dataset needs to include balanced and varied data), this student has recognized the potential harms of training AI models with limited data and the need of constructing a large dataset. In the exit interview, many students expressed their ideas of minimizing potential bias of an AI technology they are going to build, e.g., "*First I will get all types of things of specific systems. For example, to define what color crayon is, I wouldn't just get like red crayons. I would get all different types, like the rainbow and even more because there's a lot of different types of colors. And I would double check. I would keep on checking over it. I would check it with every type. So after I made it, I would just pick random crayons and see if it got the color right. And if it didn't, I would train the data set on that specific color to give it more information.*"

These findings suggest that our approach of interweaving ethics education with learning of technical concepts is highly promising in terms of preparing youth to become responsible and mindful consumers and future developers of AI technologies. They also demonstrate the success of the trajectory of learning ethics, which starts with first stimulating students' ethical imagination, then engages them in recognizing and analyzing ethical issues, and finally improves students' ethical judgment and willpower through a design project. Overall this work contributes to the AI and the design education field by providing a working learning trajectory for teaching ethics among middle schoolers. It also reinforces the importance of addressing emotional engagement, intellectual engagement, and particular knowledge in ethics education.

References

[1] J. Buolamwini and T. Gebru, “Gender shades: Intersectional accuracy disparities in commercial gender classification,” presented at the Conference on fairness, accountability and transparency, 2018, pp. 77–91.
[2] K. Kirkpatrick, “Battling algorithmic bias: How do we ensure algorithms treat us fairly?,” *Commun. ACM*, vol. 59, no. 10, Art. no. 10, 2016.
[3] A. D. Selbst, “Disparate impact in big data policing,” *Ga Rev*, vol. 52, p. 109, 2017.
[4] B. P. O’Connor, “Identity development and perceived parental behavior as sources of adolescent egocentrism,” *J. Youth Adolesc.*, vol. 24, no. 2, pp. 205–227, 1995.
[5] D. Touretzky, C. Gardner-McCune, F. Martin, and D. Seehorn, “Envisioning AI for K-12: What should every child know about AI?,” presented at the Proceedings of the AAAI Conference on Artificial Intelligence, 2019, vol. 33, pp. 9795–9799.
[6] D. Long and B. Magerko, “What is AI Literacy? Competencies and Design Considerations,” presented at the Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–16.
[7] J. A. Boss, “The effect of community service work on the moral development of college ethics students,” *J. Moral Educ.*, vol. 23, no. 2, pp. 183–198, 1994.
[8] H. Gardner, “The tensions between education and development,” *J. Moral Educ.*, vol. 20, no. 2, pp. 113–125, 1991.
[9] J. M. DuBois and J. Burkemper, “Ethics education in US medical schools: a study of syllabi,” *Acad. Med.*, vol. 77, no. 5, pp. 432–437, 2002.
[10] G. Baxter and I. Sommerville, “Socio-technical systems: From design methods to systems engineering,” *Interact. Comput.*, vol. 23, no. 1, pp. 4–17, 2011.
[11] P. Kroes, M. Franssen, I. van de Poel, and M. Ottens, “Treating socio-technical systems as engineering systems: some conceptual problems,” *Syst. Res. Behav. Sci. Off. J. Int. Fed. Syst. Res.*, vol. 23, no. 6, pp. 803–814, 2006.
[12] C. E. Harris Jr, M. Davis, M. S. Pritchard, and M. J. Rabins, “Engineering ethics: what? why? how? and when?,” *J. Eng. Educ.*, vol. 85, no. 2, pp. 93–96, 1996.
[13] B. Newberry, “The dilemma of ethics in engineering education,” *Sci. Eng. Ethics*, vol. 10, pp. 343–351, 2004.