Teaching Ethics
Online First: July 23, 2025
doi: 10.5840/tej2025722164

Teaching the Ethics of AI and Robotics to Graduate Students: A Cross-Disciplinary and Cross-Cultural Approach

Beyza Nur Guler
Virginia Tech

Yixiang Shawn Sun
Virginia Tech

Kylee Shiekh
Virginia Tech

Yi Cao
University of North Carolina at Charlotte

Michelle Ausman
Virginia Tech

Vladimir Sanchez Padilla
Virginia Tech and Universidad ECOTEC

Talha Bin Asad
Virginia Tech

Qin Zhu
Virginia Tech

ABSTRACT: As artificial intelligence and robotics are increasingly integrated in graduate research and education, graduate students across disciplines need to develop a "technological literacy" in how they work along with the ethical understanding needed to navigate these technologies responsibly. To satisfy this need, the corresponding and last author has developed a graduate-level course on AI ethics and human-robot interaction (HRI) designed for students from a variety of disciplines and backgrounds. The paper offers an overview of the course, detailing its content, institutional context, and the rationale behind its development. It describes the curriculum structure, including key themes and learning objectives, and the pedagogical approaches and assessment methods utilized in the course. The paper concludes with reflections from the instructor

on the lessons learned from teaching the course and the experiences gained throughout the learning process.

KEYWORDS: Artificial Intelligence; Human-Robot Interaction; Ethics Education; Computing Education; Engineering Education

Introduction

EMERGING TECHNOLOGIES, SUCH AS ARTIFICIAL INTELLIGENCE (AI) and robotics, have ushered in a wealth of opportunities alongside significant challenges across many aspects of human life. Graduate students from diverse disciplinary backgrounds such as Computer Science, Chemistry, and Education have endeavored to investigate how AI and robotics could potentially transform their respective fields (Tian, et al. 2024). These students encounter dual challenges posed by advancements in AI and robotics. On the one hand, they are grappling with the question of whether and how traditional theories, methodologies, and tools in their own fields can be augmented by AI-enabled technologies. On the other hand, they are expected to investigate how the natural, social, and engineering phenomena may potentially be mediated by AI-enabled technologies. We argue that familiarity with AI and robotics could potentially become a form of technological literacy that is essential for graduate students, irrespective of their academic backgrounds. Building upon such a premise, this paper introduces a graduate-level course on the ethics of AI and human-robot interaction (HRI), developed by the corresponding and last author for students from diverse cultural and disciplinary backgrounds.

This paper has four aims. Firstly, this paper provides an overview of the course, offering a general description of its contents, the institutional context in which it was conceptualized and developed, as well as the justification and rationale behind its design. Secondly, this paper introduces the curriculum structure of this course, including major themes and learning outcomes. The curriculum structure has two distinct characteristics: (1) Given the *interdisciplinary* nature of AI and robotics research, this course integrates themes and materials from a wide range of fields, ranging from philosophy, psychology, policy, HRI, to computer science. It includes philosophical literature and empirical studies. (2) Given the *global* nature of AI and robotics development and deployment, this course incorporates perspectives from both Western and non-Western cultures (e.g., Confucianism, Buddhism, African cultures). Thirdly, this paper introduces the pedagogical strategies and associated assessment methods adopted in this course. More specifically, it highlights some engaged pedagogies employed to facilitate students' learning, such as class presentations, movie reflections, AI-enhanced learning reflections, and final research papers. For the AI-enhanced learning reflections, students were asked to freely employ any generative AI tools such as ChatGPT and Bard (now Gemini) to serve as their learning companions.

This paper will summarize the opportunities and ethical concerns arising from using these tools, as identified in students' reflections. Finally, this paper will encompass reflections from the instructor regarding the lessons learned throughout the teaching and learning experiences within this course.

We hope that these reflections will offer insights aimed at further enhancing the overall learning experiences within this course. We also hope that our paper can contribute to broader discussions concerning (1) the systematic and effective design of AI ethics interventions tailored to students with diverse backgrounds, learning styles or expectations, and (2) the conceptualization and evaluation of the processes involved in designing such AI ethics interventions.

Institutional Context

Like many other institutions across the United States and worldwide, Virginia Tech has identified AI as a key strategic area for both teaching and research endeavors. Of the four research frontiers established by Virginia Tech, AI occupies the foremost position, serving as both the primary frontier and a vital enabler for advancements in the health, security, and quantum domains. At Virginia Tech, ethics and social responsibility are integral to AI research, as reflected in the university's vision for the AI research frontier:

> Building on expertise in artificial intelligence, data science, systems engineering, neuroscience, human factors, robotics, immersive visualization, and education, among others, [the AI research frontier aims] to accelerate human-technology partnerships toward seamless augmentation, ethically and sustainably. (Virginia Tech Research and Innovation n.d.)

To systematically promote AI research and education on campus, Virginia Tech has established buildings, programs, research centers, and more recently a new campus in the DC area devoted to AI and computing related research. Other interdisciplinary, AI-related research programs and institutes at Virginia Tech include the Sanghani Center for Artificial Intelligence and Data Analytics, the Hume Center for National Security and Technology, and the Tech for Humanity Initiative. Through strategic collaboration with Amazon and support from Virginia's $1 billion investment in the tech pipeline initiative, Virginia Tech aims to double its existing computer science and computer engineering programs in Blacksburg and launch the Virginia Tech Innovation Campus in National Landing. With a rapidly growing population of graduate students in AI-related fields on both the main campus and the Innovation Campus, there is a clear and timely need for courses that critically engage with the ethical, societal, and cultural dimensions of AI and robotics.

Virginia Tech has demonstrated a strong institutional commitment to cultivating AI literacy across its community. The formation of the AI Working Group in 2024, initiated by university leadership, underscores this priority, as the group is tasked with developing a systematic framework for the ethical and responsible

use of AI. This includes raising awareness, assessing risks and benefits, and guiding faculty, staff, and students. In this context, a graduate course on the ethics of AI and robotics directly supports Virginia Tech's mission by equipping graduate students—future researchers and professionals—with the critical tools needed to responsibly navigate and shape the rapidly evolving AI landscape.

More specifically, to ensure academic integrity and responsible use of emerging technologies in education, Virginia Tech has formulated specific guidelines for the use of AI-powered writing and research technologies, such as ChatGPT. These policy documents or resources aim to provide more guidance about how to use ChatGPT for teaching and research, in courses and programs at both the undergraduate and graduate levels. Virginia Tech's technology-enhanced learning and online strategies (TELOS) center also initiated AI and ChatGPT related policy and guidance in teaching.

These institutional efforts to promote AI research and education aim to foster a multidisciplinary understanding of how AI technologies can be developed and utilized to augment human capabilities ethically and sustainably. Virginia Tech is known for its strength in engineering education and the social studies of science, technology, and engineering. In addition, it hosts one of the pioneering programs in science, technology, and society (STS), which has given rise to a subfield known as Engineering Studies within STS. In general, science, engineering, business, and humanities graduate programs at Virginia Tech all have a strong emphasis on AI and other emerging technologies. Within the framework of such an institutional context, this paper describes the development of a graduate course centered on the ethics of AI and HRI, including its curriculum structure, pedagogical methodologies, and assessment strategies. Our goal is to equip students with the basic knowledge and skills necessary to explore the complex ethical landscape of AI and robotics in their everyday research practices.

Foundations of AI and Robot Ethics Education: Background Insights

With the evolving nature of AI, ethical considerations surrounding its applications have become increasingly important. As a result, AI-related courses, as well as their ethical dimensions, have been integrated into the higher education curriculum when preparing for the future AI workforce. In this case, understanding what constitutes "ethical AI" and the necessary ethical requirements, technical standards, and best practices for its realization has become a crucial issue (Jobin et al., 2019; Lauer, 2021). At least two curricular approaches toward AI ethics education have been identified: (1) offering standalone AI ethics courses, and (2) integrating ethics seamlessly into technical AI courses (Garrett, 2020). The former, whose targeted students are not limited to STEM students, often cover more diverse topics; as in Garrett et al. 2020's case, eight topics are identified. The latter, given their technical focus, tend to focus less on ethical theories, mostly covering bias, fairness, and privacy. Moreover, broadly speaking, the principles of Fairness, Accountability, Transparency, and Ethics (FATE) have been

embedded within the AI curriculum (Bogina et al., 2022). We also see some other pedagogical materials, examples include Princeton's publication of AI ethics case studies (The Princeton Dialogues on AI and Ethics, 2018). However, if we imagine "ethics" as possessing "powerful teeth," then the fragmented and inconsistent approaches to AI ethics have contributed to what Resseguier and Rodrigues (2020) describe as a "toothless" AI curriculum—highlighting the urgent need for a more proactive and integrated inclusion of ethical considerations in current curriculum (see also Borenstein and Howard, 2021).

The problems with unsophisticated integration of AI ethics can be illuminated by considering an analysis of pedagogy in the discipline of computer science. A survey of AI ethics curricula shows that the majority of ethics courses are being taught from within the discipline (Fiesler et al. 2020), even though the discipline is commonly criticized for its lack of ethical engagement. Eden has identified three broad paradigms of teaching computer science: technocratic, rationalist, and scientific (Eden 2007 135–167). The technocratic paradigm centers on the skills to build computer programs from an engineering or programmatic approach; the rationalist paradigm is a mathematically theoretical approach focusing on the underlying mathematical reality of computer systems; and the scientific paradigm is characterized by its focus on empiricism, seeking to deeply understand the behaviors of computer programs. These three paradigms shape how computer science operates, through enculturation in computer science education and in industry cultures (Raji et al. 2021). Scholars have noticed an imbalance between the "technical" and the "social"—the "hard" and the "soft"—which prioritizes the former (Raji et al. 2021). And Raji et al. (2021) argue that ethics modules developed for and within computer science are not sufficient intervention to teach CS students of how ethical challenges should be identified and resolved in real world contexts, as most require interdisciplinary discourse.

In 2020, AI ethics—including courses on security, safety, accountability, and explainability—accounts for 14% of the CS curriculum on average, while AI applications—such as courses on big data, the internet of things, and virtual reality—take a similar share on average (Zhang et al. 2021). In 2021, Raji et al. examined 254 AI ethics courses—151 explicitly marked as undergraduate courses and 92 marked as explicitly graduate—from 132 universities. Only 84 courses had full syllabi available to researchers to examine, 58 undergraduate courses and 26 graduate courses (Raji et al. 2021). Several courses did not indicate explicitly the need for cross-disciplinary collaboration, and most did not provide tools or strategies to educate students on how to seek and navigate such collaborations (Raji et al. 2021). None of the courses surveyed explicitly incorporated a critical examination of the limitations inherent in their own disciplinary perspectives. Most were designed for relatively homogenous student cohorts and included prerequisites that effectively restricted participation to students from specific academic backgrounds.

In the years since Raji et al.'s analysis, several interdisciplinary graduate courses, certificates, and degrees have been created that are free of such shortcomings. Princeton's A.I. Ethics Cohort for graduate students from various disciplines including computer scientists, engineers, sociologists, psychologists, legal scholars and others, is a notable example. Through guest speakers from diverse backgrounds, such as the Center for Human Values and the Department of Computer Science, the cohort investigates the ethical implications of AI applications in diverse fields. Purdue's M.S. in Artificial intelligence requires an ethics course and offers an AI Management and Policy focus. The University of Cambridge also offers an M.S. in AI Ethics and Society. What remains unclear is the level of interdisciplinary collaboration, student engagement, disciplinary perspectives and whether students are developing through the courses and degrees.

Despite these promising developments, critiques of AI ethics education remain ongoing. Some argue that AI ethics principles or education are deemed "useless" or relegated to the periphery, taught only when "time allows" (Garrett et al., 2020). Munn (2023) also laments the failure to effectively address the racial, social, and environmental implications of AI technologies, urging a broader consideration of systems of oppression and a narrower focus on accuracy and auditing. In the same vein, Raji et al. (2021) critique the prevailing AI ethics education paradigm, characterizing it as "exclusionary pedagogy," advocating instead for a collaborative, holistic, and ethically generative approach in AI education. Echoing this, for example, Zembylas (2023) advocates for a decolonial approach to AI in higher education teaching and learning, emphasizing the need to overcome conventional paradigms that oftentimes neglect digital neocolonialism with the advent of AI.

Slavkovik (2020) describes the challenges of teaching a course in 2020 on AI ethics to M.S. students in information science. They deemed no textbook acceptable, so the master's students struggled to read philosophical literature and scientific articles on AI, which required skills many lacked. They also describe how focus of the training of students in technical courses is their ability to "think in an algorithmic way," to see structured data in knowledge, to transform a problem into an algorithm, break it down into smaller programs and, from there, identify programming instructions (Slavkovik 2020). In mathematics and computer science solutions are often correct or incorrect. Students are "trained" to check for correctness, which often causes issues in dealing with ethical and philosophical problems that may not have an "objectively" correct solution. This training presented a bigger challenge for the students than Slavkovik anticipated.

Curriculum Structure

This section describes the curriculum structure of Virginia Tech's course on "The Ethics of AI and Human-Robot Interaction," including major themes and learning outcomes. The curriculum is distinctive in two ways: (1) It integrates interdisciplinary content—from philosophy and psychology to policy, HRI, and

computer science—combining conceptual and empirical materials; and (2) it incorporates both Western and non-Western cultural perspectives, reflecting the global nature of AI and robotics.

More specifically, this course aims to help graduate students achieve the following learning outcomes. After completing this course, students are expected to:

1. understand the basic ethical theories, concepts, tools, and frameworks for analyzing the social and ethical ramifications of AI and robotics;
2. be able to critically examine the ethical significance of the use of AI and robotics in daily and technical fields including machine learning, HRI, medicine, relationship, military, etc.
3. understand laws, policies, and politics related to the governance of AI and robotic technologies;
4. develop a critical attitude toward the role of AI and robotics in shaping human society including human perceptions and behaviors;
5. be able to use the theories, concepts, tools, and frameworks learned from this class to criticize emerging AI and robot ethics issues in the society.

Rather than using a textbook, the assigned content consisted of peer-reviewed research papers, case studies, websites, and AI- and robot-related videos on the implications of AI and robotics along with human interactions. Students were expected to contribute additional references needed to complete their assignments and presentations throughout the semester. The instructor adapted the curriculum to align with students' specific interests and deepen their understanding of this evolving field. It allowed the integration of diverse resources to promote participation through different perspectives and enhance critical thinking skills. The course avoided relying on textbooks due to the evolving nature of the field; instead, it drew on current literature and diverse resources to provide in-depth insights into the societal and cultural implications of AI and robotics.

This course builds on a "Robot Ethics" course the corresponding author had co-taught and taught at Colorado School of Mines. One major reason for the name change from "Robot Ethics" to "The Ethics of AI and Human-Robot Interaction" was to reflect the interrelated nature of the problem. The advancement of generative AI tools is bolstering robotics, forging a closer bond between intelligence and robotics. The science of AI grapples with inquiries regarding knowledge, reasoning, and the manipulation of knowledge, while robotics, particularly in human-robot interaction, presents a challenge by bridging intelligence with tangible action and real-world objects. Hence, conventional subjects within robot ethics or the ethics of HCI frequently encompass social and ethical dilemmas stemming from the *tangible presence* of robots. These encompass considerations such as their physical appearance, linguistic capabilities, and interactions with

humans. In contrast, AI ethics focuses on the ethical and political implications of intelligent systems, algorithms, and machine learning, addressing issues like bias, discrimination, and other forms of injustice. Therefore, AI ethics issues are also intertwined with broader policy and governance of AI. The curriculum structure was deliberately designed to encompass both concrete issues related to the ethics of HRI research and innovation and broader concerns related to the ethics and policy of AI.

Table 1 compares the topics in this course under the domains of AI ethics and the ethics of HRI, noting that some topics overlap between the two domains:

Table 1. Curriculum Topics in the Course

AI Ethics Topics	**Ethics of HCI Topics**	**Cross-domain Topics**
AI in non-Western philosophical traditions such as Confucianism, Buddhism, and African philosophy AI, environmental ethics, and global climate change race, gender, and AI the use of AI in spirituality AI and political philosophy global perspectives of AI governance (e.g., Europe, East Asian, and Middle East) future of work including crowdwork professional norms for AI practices	engineering moral reasoner (e.g., how to design robots capable of reasoning morally like humans) robot rights trust and deception in HRI anthropomorphism robot friends humanoid sex and love robots moral persuasion (e.g., how robotic moral advice affects human moral behavior)	moral theories for designing AI and robots artificial moral agency the use of AI and robots in education

As indicated earlier, one of the two features of this course is its interdisciplinarity. It draws on resources and theories from a wide range of fields, ranging from philosophy, psychology, policy, HRI, to computer science. More specifically, it addresses not only "ethics for AI and HRI" (ethical regulations for AI and HRI) but also "AI and HRI for ethics" (using AI and HRI innovations to achieve more socially desirable outcomes). As indicated in the course syllabus, this course investigates

- the social and ethical implications of AI and robotics;
- empirically informed efforts to design human-robot interaction that generates positive moral outcomes.

As indicated in Table 1, most issues are related to the *ethics for AI and HRI*. Some issues in Table 1 specifically address *AI and HRI for ethics*, or the role of AI and HRI in ethical analysis and innovation. For instance, under the theme of moral persuasion of robots, the instructor included empirical studies from his collaboration with computer scientists and social psychologists. These studies focus on the design and assessment of how wakewords[1] and robotic moral advice influence human moral behavior, reflection, and perception.

Besides its interdisciplinary nature, another feature of the curriculum is that it covers both philosophical and policy perspectives from non-Western traditions. For instance, it includes philosophical discussions of AI and robotics from Confucian, Buddhist, and African perspectives. The inclusion of these articles aims to demonstrate to students that there are alternative approaches to examining the personhood, agency, and rights of AI systems and robots. These approaches contrast with the dominant Western properties perspective, which argues that intrinsic metaphysical properties determine the personhood of AI systems and robots (Gunkel 2023). From the non-Western perspectives, the personhood of AI and robots is determined by their relationships to us, the social roles they assume in the community, and the extent to which they exhibit behaviors that would qualify them to be included in the community *by humans* (Hongladarom 2020, Wareham 2021, Zhu, et al. 2020). From the policy perspective, this course also included readings from the *Oxford Handbook of Ethics of AI* that examine the perspectives, approaches, and governance frameworks for AI policy in the Global South including Asia and Middle East. There are at least several strengths associated with including these non-Western resources:

- they allow for the comparison and contrast of non-Western resources with Western-centric literature, enabling an examination of the fundamental assumptions and limitations of Western approaches;
- they engage international students by encouraging them to bring their own cultural perspectives and experiences to the classroom and contributing to class discussions, which is particularly valuable given that they constitute a significant portion of graduate students;
- they help graduate students, especially those in AI-related fields, to formulate AI design solutions that are responsive to users with culturally diverse backgrounds.

Pedagogies for AI and Robot Ethics Education

This section introduces the pedagogical strategies adopted in this course. More specifically, it highlights some engaged pedagogies employed to facilitate students' learning, such as class presentations, movie reflections, AI-enhanced learning reflections, and final research papers. For the AI-enhanced learning reflections, students were asked to freely employ any generative AI tools such as ChatGPT and Bard (now Gemini) to serve as their learning companions. Later

in this section, we summarize the opportunities and ethical concerns identified in students' reflections on the use of these tools. Overall, this course incorporated five pedagogical strategies:

(1) Weekly Online Discussion Posts

Students engaged in Canvas discussion boards by reflecting on what they learned from assigned weekly readings and responding to at least two comments posted by their peers.

(2) Group Facilitation and Discussion

Each student facilitated class discussions on weekly assigned readings and topics twice. Facilitations involved presentations and interactive classroom discourse.

(3) AI-Enhanced Learning Reflections

In this reflective writing assignment, students engaged in critical reflection regarding their weekly use of ChatGPT and/or other Generative AI tools as a learning companion. The assignment prompted students to compose a paragraph reflecting on their experiences with the tools while working on various tasks including but not limited to course—related activities such as reading and writing assignments. Students were encouraged to consider various aspects of their interaction with these tools, such as the purpose of engagement, its impact on intellectual thinking, writing and communication skills, as well as its influence on personal or scholarly growth. Furthermore, students were prompted to consider potential ethical concerns surrounding research or the societal role of ChatGPT in society.

The submission format was flexible allowing students to choose between paragraph or bullet point formats. Additionally, students could supplement their reflections with attached notes or screenshots of their conversations with ChatGPT.

The following two tables summarize the opportunities and challenges identified in students' responses:

Table 2. Opportunities Identified in Students' Responses in Using AI as a Learning Companion

Opportunities	
Theme	**Description**
Brainstorming Support	Highlights ChatGPT's assistance in facilitating brainstorming by generating ideas and/or general information as a starting point for writing.
Organizing Writing and Outlining	Describes ChatGPT's ability to assist in organizing writing by generating outlines and summarizing or condensing information.
Nuances of Generative AI	Refers to the differences between different generative AI tools such as Bard, Gemini, ChatGPT, Consensus and Grammarly. This code involves understanding the context each tool is useful for and accordingly their strengths and limitations.
Grammar and Editing Assistance	Indicates ChatGPT's utility in assisting with grammar and editing tasks including proofreading and improving writing quality.

Table 3. Challenges Identified in Students' Responses in Using AI as a Learning Companion

Challenges	
Theme	**Description**
Limited Coverage of Information	Highlights issues regarding the depth and specificity of ChatGPT's responses. Since ChatGPT's responses are restricted to information prior to 2022, the information provided is not up to date.
Reliability and Accuracy	Indicates concerns about the trustworthiness and precision of ChatGPT's responses
Cautious Use and Crosschecking	Reflects the necessity to double check the output of ChatGPT due to the lack of source attribution.
Accessibility and User Experience	Demonstrates the different levels of depth obtained by different Generative AI tools for the same prompt. This also includes instances where students mention the need for a specific audience and tone to receive the desired output.
Lack of Advanced Writing Ideas and Source Attribution	Describes ChatGPT's inability to generate advanced writing ideas (i.e., dissertation abstract)

(4) Movie Reflections

Throughout the whole class, students were asked to write four movie reflection essays. In the movie reflection assignment, students must write a two-page, double spaced reflective essay after watching a video clip or movie. The essay should establish connections between the theories covered in class and the content of the movie. Each reflection assignment included guiding questions or writing prompts. These prompts explored topics covered in the readings and/or class discussions such as the moral agency of robots, human perception of robots and the societal implications of co-existing with robots.

The movie clips chosen for the assignments were mostly from the "Robot Film Festival (https://robot.film/)." The Robot Film Festival, a one-of-a-kind yearly event from 2011 to 2018, sought to bring creativity and fun to traditional robotic engineering while exploring the growing connections between humans and machines. It featured a wide range of films, including invited works and selected submissions, along with live performances by robot entertainers and an exciting red-carpet awards ceremony. These robot-themed films typically run around 5 minutes in length, with robots taking center stage as the primary characters. Most these movies cover the topics typical in the AI and robot ethics literature. Table 4 provides an example of a movie reflection assignment, along with accompanying writing prompts or guiding questions.

Table 4. An Example for the Movie Reflection Essay Assignments

Watch the video "No Robots" (https://vimeo.com/groups/robotfilmfest/videos/23017365) from the Robot Film Festival and draft a two-page, double-spaced reflection essay. The assignment prompts you to connect theories learned in class with the movie you'll watch. Here are some guiding questions to consider while working on the assignment, though feel free to include additional questions that interest you: • Should the robot featured in the movie be entitled to certain moral rights? What justifies such entitlement, or lack thereof? Can these considerations be aligned with any theories discussed in our course? • What concerns do humans have about robots in their society and everyday life? • What changes have you discovered in terms of the moral perceptions, behaviors, or decision-making of the man in the video? • What are some typical social images of the robot you find in the video? To what extent you find these images convincing or not? Where do these images come from? • What are some implications you can draw from this video for better understanding and living with robots?

5) Final Research Paper

Students were tasked with completing a research paper as their final project. They were encouraged to explore the reciprocal relationship between AI or robotic technology and human society, examining how each shapes the other. Throughout the latter portion of the course, the instructor collaborated individually with students on progressive assignments, including selecting a topic,

crafting an extended abstract, outlining the paper, and completing the final draft. The instructor also provided a general outline for the paper that included the following sections:

Motivation and Justification: Why is this research important for fields of AI and robot ethics, human-robot interaction, and/or your own field of interest such as engineering education? Any potential gaps in the literature you want to address?

Background: This could be a literature review section or something similar. It does not need to be a systematic review. Rather the goal can be quite pragmatic. For instance, what is the information needed for non-experts to understand the context of your study? What are the most important papers/studies in the literature? What are the articles that will help justify your study design and/or methods? Any other sources helpful for further justify the value of your research?

Methods: What are the techniques and procedures you will need to collect your data? What methodological approach(es) will this study take? It could be a conceptual study, a more traditional empirical study, a pilot study for a much larger project, a review paper, a policy essay, etc. No matter what approach you are taking, make sure you state it clearly in the paper.

Findings: You are expected to systematically present the insights that are generated through your inquiry. You have the freedom to determine the appropriate way(s) to present your findings. You could either mainly synthetize your findings or combine your presentation of findings and engagement with the relevant literature. Please note visual aids are always useful for organizing findings.

Discussions: Based on what you have presented in the Findings section, now you are expected to discuss much broader implications of your study. How are these findings relevant to other scholars interested in this topic or in your field? What future directions are you anticipating?

Given all students came from the field of engineering education, the topics of their final projects were all related to either the ethical implications of using AI in higher education or the formation of professional engineers for AI workforce. For instance, their research topics included:

- Investigating the perceptions of graduate students in AI-related fields regarding the environmental impacts and sustainability of AI;
- Exploring the perspectives of AI scientists and engineers in policy-related roles regarding their responsibilities in AI development and deployment;
- Understanding the decision-making process of graduate students regarding the integration of AI-enabled tools into their research endeavors.

Reflections

This course was offered as a three-credit hour special topic class within the Department of Engineering Education at Virginia Tech, reflecting the academic affiliation of the instructor. Because this class was categorized as a "special topic" and many graduate students across campus were unfamiliar with course offerings in the Department of Engineering Education, the instructor encountered unexpected challenges in recruiting students. Students interested in enrolling in a course on AI and robot ethics might anticipate finding it within humanities departments like STS or Philosophy, or within technical departments closely linked to AI, such as Computer Science or Industrial Engineering. Another complicating factor in teaching and promoting this course is that graduate curriculum requirements typically do not require credit hours in general education, unlike the undergraduate curriculum. Hence, drawing from the findings provided in the "Background Insights" section, an effective approach to promote such a course could involve integrating it into a graduate program, a minor program, or a certificate program related to AI. However, implementing such a strategy requires substantial effort in coordinating and negotiating with faculty members within these programs which often can be daunting. Certainly, another possibility could be collaborating with faculty who are teaching AI-related courses and designing ethics modules and integrating them into technical courses. However, clear motivations or mechanisms are necessary to encourage interdisciplinary collaboration among faculty members from engineering, computer science, humanities, and social sciences.

Given the advanced level of this AI and robot ethics course for graduate students, defining clear learning expectations and outcomes poses a challenge. Unlike undergraduate students, who typically focus on coursework with occasional research projects, graduate students, particularly those in science or engineering programs, primarily dedicate their time to conducting research within their respective fields. Should the course learning outcomes prioritize providing a liberal arts or general education perspective for students, or should they emphasize offering interdisciplinary research tools and opportunities directly beneficial to their research projects or activities? The curriculum design of this course aligns with a framework that encompasses both approaches to learning outcomes in AI ethics education: (1) ethics for AI and HRI; and (2) AI and HRI for ethics. Such a framework can be potentially useful for students and faculty to clarify and justify the learning outcomes of their own AI ethics courses.

Another challenge noted by the instructor while teaching this course was that students frequently encountered difficulty in understanding the assigned reading materials for several reasons: (1) there are no textbooks tailored specifically for graduate-level students; (2) the assigned readings were often written for experts in the respective fields rather than a general audience; and (3) students lacked training in related disciplines such as ethics, psychology, and other social sciences, which hindered their ability to engage the materials effectively. These

challenges further impede students' capacity to apply concepts and theories from the readings and class discussions to their assignments effectively.

With an increasing number of graduate degree and certificate programs in AI-related fields both nationally and globally, the development of effective AI ethics education programs becomes imperative. Hence, a prospective research project for engineering and computing education researchers could be a nationwide project aimed at understanding the needs of students and faculty in AI graduate programs, as well as the expectations of employers hiring these graduates. Such an initiative would not only facilitate improved student engagement and learning outcomes but also enhance their readiness for future career prospects in AI research and innovation.

Acknowledgments

This material is based upon work supported by the National Science Foundation under grant numbers: 2124984, 2412398, 2418848, and 2418867. It was also partially funded by the Air Force Office of Scientific Research (AFOSR) project "Developing a Quantification System for Robot Moral Agency." Special thanks to Dr. Glen Miller and Dr. Deborah Mower for their excellent editing work, which greatly enhanced the clarity and quality of this material.

Note

1. "Wakewords" are specific phrases or words that trigger a voice-controlled device, such as a smart speaker or virtual assistant, to activate and start listening for further commands or instructions. For example, "Hey Siri" or "Okay Google" are wakewords used to activate Apple's Siri and Google's virtual assistant, respectively. These wakewords are designed to prompt the device to start processing voice input and perform tasks or provide information based on the user's voice commands.

References

Bogina, Veronika, Alan Hartman, Tsvi Kuflik, and Avital Shulner-Tal. 2021. "Educating Software and AI Stakeholders about Algorithmic Fairness, Accountability, Transparency and Ethics." *International Journal of Artificial Intelligence in Education* 32 (April). https://doi.org/10.1007/s40593-021-00248-0

Borenstein, Jason, and Ayanna Howard. 2020. "Emerging Challenges in AI and the Need for AI Ethics Education." *AI and Ethics* 1 (1). https://doi.org/10.1007/s43681-020-00002-7

Eden, Amnon H. 2007. "Three Paradigms of Computer Science." *Minds and Machines* 17 (2): 135–67. https://doi.org/10.1007/s11023-007-9060-8

Fiesler, Casey, Natalie Garrett, and Nathan Beard. 2020. "What Do We Teach When We Teach Tech Ethics?: A Syllabi Analysis." In *SIGCSE '20: Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. https://doi.org/10.1145/3328778.3366825

Garrett, Natalie, Nathan Beard, and Casey Fiesler. 2020. "More than 'If Time Allows': The Role of Ethics in AI Education." In *AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. https://doi.org/10.1145/3375627.3375868

Gunkel, David J. 2023. *Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond*. Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/14983.001.0001

Hongladarom, Soraj. 2020. *The Ethics of AI and Robotics: A Buddhist Viewpoint*. Lanham, MD: Lexington Books. https://doi.org/10.5771/9781498597302

Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1 (9): 389–99. https://doi.org/10.1038/s42256-019-0088-2

Lauer, Dave. 2020. "You Cannot Have AI Ethics without Ethics." *AI and Ethics* 1 (October). https://doi.org/10.1007/s43681-020-00013-4

Marija Slavkovik. 2020. "Teaching AI Ethics: Observations and Challenges." *Norsk IKT-Konferanse for Forskning Og Utdanning*, no. 4 (November).

Munn, Luke. 2022. "The Uselessness of AI Ethics." *AI and Ethics* 3 (3). https://doi.org/10.1007/s43681-022-00209-w

Raji, Inioluwa Deborah, Morgan Klaus Scheuerma, and Razvan Amironesei. 2021. "You Can't Sit with Us: Exclusionary Pedagogy in Ai Ethics Education." In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3442188.3445914

Rességuier, Anaïs, and Rowena Rodrigues. 2020. "AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics." *Big Data & Society* 7 (2). https://doi.org/10.1177/2053951720942541

Tian, Weiqi, Jingshen Ge, Yu Zhao, and Xu Zheng. 2024. "AI Chatbots in Chinese Higher Education: Adoption, Perception, and Influence among Graduate Students—An Integrated Analysis Utilizing UTAUT and ECM Models." *Frontiers in Psychology 15*. https://doi.org/10.3389/fpsyg.2024.1268549

Virginia Tech Research and Innovation. n.d. *The Artificial Intelligence Frontier. Accessed May 18, 2024.* https://www.research.vt.edu/initiatives/research-frontiers/artificial-intelligence.html.

Wareham, C. S. 2021. "Artificial Intelligence and African Conceptions of Personhood." *Ethics and Information Technology* 23: 127–136. https://doi.org/10.1007/s10676-020-09541-3

Zembylas, Michalinos. 2021. "A Decolonial Approach to AI in Higher Education Teaching and Learning: Strategies for Undoing the Ethics of Digital Neocolonialism." *Learning, Media and Technology* 48 (1): 1–13. https://doi.org/10.1080/17439884.2021.2010094

Zhang, Daniel, Saurabh Mishra, Erik Brynjolfsson, John Etchemendy, Deep Ganguli, Barbara Grosz, Terah Lyons, et al. 2021. "The AI Index 2021 Annual Report." arXiv. http://arxiv.org/abs/2103.06312.

Zhu, Qin, Tom Williams, Blake Jackson, and Ruchen Wen. 2020. "Blame-Laden Moral Rebukes and the Morally Competent Robot: A Confucian Ethical Perspective." *Science and Engineering Ethics* 26: 2511–2526. https://doi.org/10.1007/s11948-020-00246-w