RESEARCH   **Open Access**

Unpacking the role of AI ethics online education for science and engineering students

Maya Usher[1]* and Miri Barak[1]

Abstract

Background As artificial intelligence (AI) technology rapidly advances, it becomes imperative to equip students with tools to navigate through the many intricate ethical considerations surrounding its development and use. Despite growing recognition of this necessity, the integration of AI ethics into higher education curricula remains limited. This paucity highlights an urgent need for comprehensive ethics education initiatives in AI, particularly for science and engineering students who are at the forefront of these innovations. Hence, this research investigates the role of an online explicit-reflective learning module in fostering science and engineering graduate students' ethical knowledge, awareness, and problem-solving skills. The study's participants included 90 graduate students specializing in diverse science and engineering research tracks. Employing the embedded mixed-methods approach, data were collected from pre- and post-intervention questionnaires with closed-ended and open-ended questions.

Results The study's results indicate that the online explicit-reflective learning module significantly enhanced students' knowledge of AI ethics. Initially, students exhibited a medium–high level of perceived ethical awareness, which saw a modest but statistically significant enhancement following the participation. Notably, a more distinct increase was observed in students' actual awareness of ethical issues in AI, before and after the intervention. Content analysis of students' responses to the open-ended questions revealed an increase in their ability to identify and articulate concerns relating to privacy breaches, the utilization of flawed datasets, and issues of biased social representation. Moreover, while students initially displayed limited problem-solving abilities in AI ethics, a considerable enhancement in these competencies was evident post-intervention.

Conclusions The study results highlight the important role of explicit-reflective learning in preparing future professionals in science and engineering with the skills necessary for ethical decision-making. The study highlights the need for placing more emphasis not only on students' ability to identify AI-related ethical issues but also on their capacity to resolve and perhaps mitigate the impact of such ethical dilemmas.

Keywords Artificial intelligence (AI), Case-based learning, Ethics education, Ethical awareness, Higher education, Science and engineering education

Introduction

As with every rapidly growing technology, Artificial Intelligence (AI) holds both great promise and threat. It holds the promise to deliver a better quality of human life through advancements in healthcare, environmental sustainability, and transportation accessibility (Taddeo & Floridi, 2018; Zhou et al., 2020). Despite the obvious advantages, advancements in AI bring forth a host of ethical issues associated with its development and use (Bogina et al., 2022; Nam & Bai, 2023; Zhou et al., 2020). The collection, utilization, and misuse of data for AI training can expose people to unforeseen risks (Borenstein & Howard, 2021; Taddeo & Floridi, 2018). As these systems continue to advance and perform increasingly

*Correspondence:
Maya Usher
mayau@technion.ac.il
[1] Faculty of Education in Science and Technology, Technion-Israel Institute of Technology, 320003 Haifa, Israel

complex tasks, their behaviour can be difficult to monitor, validate, predict, and explain (Bogina et al., 2022; Borenstein & Howard, 2021). These dual facets of AI underscore the pressing need for a comprehensive ethical framework within the realm of higher education (Borenstein & Howard, 2021; Ouyang et al., 2022).

Despite the critical importance of embedding AI ethics within educational curricula, efforts to integrate such critical discussions into educational frameworks are still limited (Borenstein & Howard, 2021; Zawacki-Richter et al., 2019). The absence of a holistic approach to teaching AI ethics underscores the need for educational institutions to adapt, ensuring students from all disciplines are equipped with the knowledge and tools to navigate the complex ethical landscape of AI technology (Erduran, 2023; Holmes et al., 2022).

Given AI's pervasive impact, instruction on AI ethics is especially crucial for students in science and engineering fields, who are poised to lead its future development (Akgun & Greenhow, 2022; Kong et al., 2023; Usher & Hershkovitz, 2022; Xu & Ouyang, 2022). Recent studies shed light on the state of AI instruction within these disciplines, revealing a growing but uneven integration of AI and its ethical considerations (Casal-Otero et al., 2023; Park et al., 2023; Xu & Ouyang, 2022). Moreover, empirical evidence supports the efficacy of AI ethics programs in enhancing undergraduate students' understanding and awareness of ethical principles related to AI, demonstrating improvements in students' ability to identify and address ethical issues (Kong et al., 2023; Lin et al., 2023). The collective insights from these studies advocate for a more structured and critical inclusion of AI ethics in education. This is especially vital for preparing future AI developers, practitioners, and researchers with the ethical foresight required in their fields (Erduran, 2023; Xu & Ouyang, 2022).

Research goal and questions

The goal of the current study is to explore the role of an online explicit-reflective learning module in fostering ethical knowledge, awareness, and problem-solving skills of science and engineering graduate students. Three research questions guide this study:

1. Whether and to what extent does participation in the online explicit-reflective learning module enhance knowledge of AI ethics?
2. Whether and to what extent does participation in the online explicit-reflective learning module enhance both perceived and actual awareness of AI ethics?
3. Whether and to what extent does participation in the online explicit-reflective learning module improve problem-solving skills in AI ethics?

Literature review
Ethics in the field of AI

Artificial Intelligence (AI) is defined as the simulation of human intelligence in machines programmed to think and learn like humans. This technology encompasses a range of systems, from simple algorithms to complex machine learning and neural network systems, capable of performing tasks that typically require human intelligence (Zawacki-Richter et al., 2019). AI presents transformative advancements across vital sectors, including healthcare—where it can improve diagnostics and patient care, environmental sustainability—by optimizing resource use and reducing waste, and transportation—increasing accessibility and efficiency. These developments make daily life more manageable and sustainable for people around the globe (Taddeo & Floridi, 2018; Zhou et al., 2020).

Yet, alongside these advancements lurk significant ethical quandaries (Bogina et al., 2022; Nam & Bai, 2023). As AI reshapes daily practices and interactions, an ethical framework becomes paramount to harness its potential while mitigating associated risks. AI ethics, therefore, scrutinizes the moral implications of AI technologies, addressing concerns from their initial stages of development to their broader deployment and governance (Taddeo & Floridi, 2018). The scope of ethical issues related to AI is vast, encompassing a wide range of societal and ethical issues (Holmes et al., 2022; Jobin et al., 2019). First, the extensive data collection and analysis required by AI further escalate the risks of unauthorized access, data breaches, and manipulative use of private and personal information (Akgun & Greenhow, 2022; Borenstein & Howard, 2021; Jobin et al., 2019). Moreover, concerns extend to significant disruptions in the labor market driven by automation and increased productivity, alongside challenges related to intellectual property rights and authorship rights (Bogina et al., 2022; Nam & Bai, 2023; Pavlik, 2023). These challenges encompass the complexities of attributing ownership, where the contributions of humans and machines may blur traditional boundaries of intellectual property. In addition, issues of academic integrity arise as AI tools become more integrated into research and learning processes, necessitating clear guidelines to prevent misuse and ensure proper citation (Cooper, 2023; Kumar et al., 2024).

In light of AI's increasing integration into various societal aspects, it becomes imperative to provide researchers, educators, and students with the necessary tools to understand and address the complex ethical issues surrounding AI (Akgun & Greenhow, 2022; Park et al., 2023). The integration of AI ethics into educational curricula is proposed as a means to inform these key groups about AI's challenges, fostering a knowledgeable and

ethically aware community (Borenstein & Howard, 2021; Ouyang et al., 2022).

AI ethics education

Universities worldwide have long acknowledged the vital role of ethics education, dedicating substantial efforts to weave it into both undergraduate and graduate curricula (Barak & Green, 2020; Taebi & Kastenberg, 2019). These endeavors aim to embed ethical norms among future professionals, enhancing their competence to confront the ethical dilemmas inherent in their respective domains (Mitcham & Englehardt, 2019; Qiao et al., 2024). Such instruction strives to promote awareness and application of professional norms and ethical principles in the performance of scientific research (Bairaktarova & Woodcock, 2017; Qiao et al., 2024). Despite these established efforts, the specific integration of AI ethics into academic programs remains comparatively scarce (Borenstein & Howard, 2021; Casal-Otero et al., 2023; Zawacki-Richter et al., 2019).

AI ethics education plays a pivotal role in educating students about their ethical responsibilities in the development and application of AI technologies (Borenstein & Howard, 2021). This form of education is instrumental in enhancing students' critical thinking skills, preparing them to navigate the gap between ethical intentions and practical implementation in AI technologies (Holmes et al., 2022). Engagement in AI ethics education initiatives can equip students with the essential knowledge and skills necessary to tackle AI's ethical challenges effectively (Borenstein & Howard, 2021). Despite these apparent benefits, the comprehensive integration of AI ethics into educational programs remains limited (Borenstein & Howard, 2021; Zawacki-Richter et al., 2019).

The field of AI ethics education faces some key challenges, including the necessity for a multidisciplinary approach. Raji et al. (2021) criticize the current AI ethics education space for its "exclusionary pedagogy," where ethics is distilled for computational approaches without engaging with other ways of knowing, which would benefit ethical thinking (Raji et al., 2021). According to the authors, educators should focus not only on developing the technical skills or social theory skills of students. Instead, more attention should be paid to the value of appropriately articulating the right problems. Keeping pace with rapidly evolving AI technologies poses another significant challenge for AI ethics education. AI ethics curricula should be dynamic and adaptable, reflecting the latest developments and ethical challenges emerging from new AI technologies. Furthermore, ensuring the practical relevance of AI ethics education is crucial. Students should be able to apply ethical theories and principles in real-world settings, necessitating a blend of theoretical and practical learning experiences (Mouta et al., 2019).

Various pedagogical approaches have been employed in AI ethics education to reflect its evolving and interdisciplinary nature. One prevalent method is the use of case studies, which allow students to analyze ethical dilemmas encountered in real-world AI applications (Barak & Usher, 2019; Raji et al., 2021; Usher & Barak, 2020). This method aids in grounding abstract ethical concepts, making them more accessible and relatable. Recent research indicates that case studies on AI applications can significantly enhance AI literacy and ethical awareness among primary school (Lin et al., 2023) and higher education students (Kong et al., 2023).

AI ethics education is particularly crucial in science and engineering academic environments, as it has the potential to nurture a generation of professionals who are not only technically skilled but also deeply versed in the ethical complexities of AI (Barak & Green, 2020; Borenstein & Howard, 2021). Yet, only limited empirical studies explored the integration of AI ethics education in technical universities (Kong et al., 2023; Martin et al., 2021; Zawacki-Richter et al., 2019). Moreover, scant research investigated AI ethics from the perspective of graduate students who are young researchers at an early stage of their professional careers.

Methods
Research participants and setting

The study involved a diverse group of 90 graduate students specializing in various science and engineering research tracks at a leading technological university. The gender distribution among participants was 59% male and 41% female. The participant disciplines were predominantly split between Engineering, comprising 56% (with specializations such as Electrical Engineering, Computer Engineering, and Information System Engineering), and Sciences, accounting for 44% (including areas like Chemistry, Physics, and Biology). By focusing on a diverse group of graduate students across various scientific and engineering disciplines, this research underscores the cross-disciplinary importance of ethical education in the era of AI.

Upon university enrollment, graduate students across all faculties are mandatorily enrolled in a cross-disciplinary online course titled 'Ethics of Research'. Each semester, around 300 students participate in the online course. This self-paced, content-centric online course is a prerequisite for the submission of their research proposals, emphasizing individual learning from a distance. In recognition of the growing impact of AI on contemporary research, we incorporated a specialized module, 'Ethics in the Development and Use of AI', into this course.

This module is designed to equip students with the critical ethical knowledge and skills needed to proficiently address the ethical intricacies of AI technologies. Our objective is to prepare students thoroughly for the ethical challenges they may encounter in their forthcoming professional endeavors.

The learning module utilizes a blend of case-based learning with explicit-reflective exercises of about a total of five hours. It was developed based on reports outlining AI ethical guidelines, such as the OECD's Recommendation of the Council on Artificial Intelligence (2021), which provide comprehensive guidelines and ethical principles for AI development and use, and the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems' report (2019).

In addition, the module engaged students in case-based learning exercises. These exercises present students with seven real-world case studies illustrating various ethical dilemmas in AI's development and/or use. Students are asked to select one or more case studies that particularly resonate with them or spark their interest, then reflect on the ethical considerations these scenarios might entail, including potential solutions. These case studies cover compelling topics like autonomous vehicles making critical decisions about pedestrian safety, the misuse of personal data from dating websites beyond its original purpose, and the implications of AI-generated art that mimics the styles of renowned artists.

Research methods and tools

The study applied the convergent mixed methods design, incorporating both quantitative and qualitative data collection and analysis (Creswell, 2014). Data were collected through students' responses to a questionnaire that was administered at the beginning and the end of the AI ethics explicit-reflective learning module. The questionnaire included three parts, as described in the following paragraphs.

The first part focused on assessing students' knowledge of AI ethics through a series of 12 multiple-choice questions, directly addressing the first research question. This section was based on existing European reports about ethics in AI, specifically the OECD's "Recommendation of the Council on Artificial Intelligence" (OECD, 2021). Two exemplar questions are provided to illustrate the questionnaire's scope and format.

As an artificial intelligence-based system processes larger amounts of data and establishes broader connections, explaining its operations and the basis for its predictions becomes increasingly challenging. Which principle is intended to address this complexity? (Correct answer—The Principle of Explainability).

Artificial intelligence-based systems are employed in some courts to evaluate the recidivism risk of detainees. Research has shown that these systems tend to predict a higher recidivism risk for detainees belonging to specific populations compared to others. Which ethical issue does this scenario highlight? (Correct answer—Discrimination arising from the way the data is selected and the design of the algorithm).

Overall, students' scores could range from 0 (no correct answers) to 12 points (1 point for each correct answer of the 12 questions).

The second part aimed to examine students' perceived awareness of AI ethics, directly addressing the second research question. It comprised an eight-item, five-point Likert scale, ranging from 1—strongly disagree to 5—strongly agree. This part of the questionnaire was adapted from the work of Barak and Green (2021), who used it to examine students' knowledge and awareness of responsible conduct of research. Example statement: 'It is important to learn about ethical issues in AI during academic studies'; Ethics education may promote responsible behavior in AI; 'I am confident in my ability to identify ethical dilemmas related to AI'.

The scale's internal consistency, indicative of students' perceived ethical awareness, was validated through Cronbach's α coefficient, with pre-questionnaire and post-questionnaire values of 0.75 and 0.81, respectively. To ensure the reliability of results and mitigate internal validity threats, the sequence of items was varied between the pre-and post-questionnaires. Moreover, the content validity of the closed-ended items was assessed using the Content Validity Ratio (CVR) with four expert researchers in science and engineering education. Two of the experts hold PhD degrees, and the other two have master's degrees and are currently doctoral students. The four experts reached full agreement on all closed-ended items.

The third part aimed to examine students' actual awareness and problem-solving skills related to AI ethics via two open-ended questions, addressing the second and third research questions. In the current study, actual ethical awareness was examined according to students' ability to recognize an ethical dilemma or concern that may arise in a certain situation. Problem-solving skills related to AI ethics were examined by students' ability to provide potential solutions to an ethical problem.

Responses were prompted by presenting a photograph titled "A research laboratory that works on the development of artificial intelligence tools," designed to stimulate reflective thinking on potential ethical challenges and resolutions without biasing towards specific AI research areas and encourage them to contemplate research involving the development of AI from a holistic and

nuanced viewpoint. Students were instructed to carefully and critically examine the photograph and respond to two questions: a) what possible ethical issues may arise while working in this lab? b) What are the potential solutions for these ethical issues?

The study was conducted in accordance with the university's ethical guidelines and received an IRB ethical clearance. Both questionnaires (pre- and post-intervention) were fully anonymous to ensure participants' privacy and encourage honest responses. Students signed an informed consent form that detailed the goals of the research and the manner of participation. They were informed about the voluntary nature of their participation and their right to withdraw at any time without any consequences.

Data analysis

The quantitative data collected from the pre-and post-intervention questionnaires were analyzed using IBM's Statistical Program for the Social Sciences (SPSS). We performed a factor analysis to explain and interpret the results from the 12 multiple-choice questions representing students' knowledge of AI ethics. To evaluate the intervention's impact, we conducted paired-sample t tests to compare the mean scores from students' responses to the 12 multiple-choice questions (knowledge in AI ethics) and the eight-item Likert-type scale (perceived awareness of AI ethics), both before and after participation in the learning module. The magnitude of the observed differences was quantified using Cohen's d value, providing an estimate of effect size.

The qualitative data from students' responses to the open-ended questions were subjected to an in-depth analysis to trace any potential enhancement in actual ethical awareness and problem-solving skills. Utilizing a conventional (inductive) content analysis approach, as outlined by Hsieh and Shannon (2005), the analysis proceeded in four stages: first, all students' responses were compiled into a single comprehensive file. Second, the first author of this manuscript conducted a thorough examination of the responses, highlighting text segments indicative of ethical issues (demonstrating actual awareness) and potential solutions (demonstrating problem-solving skills). Third, the marked text segments were thematically categorized and numerically coded, encompassing one to ten ethical issues and one to seven potential solutions.

To ensure inter-coder reliability, a randomly selected sample of the responses, along with the established categories, were evaluated by four judges, all expert researchers in science and engineering education. Two judges with a PhD degree and long experience in the field, and two judges with a master's degree and are currently doctoral students. After each judge independently coded the responses, Cohen's Kappa analysis was performed to assess the level of agreement between their coding. The analysis confirmed high inter-coder reliability, with agreements of 90% and 94% with the first and second judges, respectively.

Each student's actual ethical awareness and problem-solving skill level were then scored on a scale from 1 to 5, with the scale defined as follows: 1 point for no response; 2 points for identifying one ethical issue or solution; 3 points for two issues or solutions; 4 points for three; and 5 points for identifying four or more ethical issues or solutions. The identified ethical issues and proposed solutions were further analyzed to detect any significant shifts post-intervention, employing the Wilcoxon signed-rank test to assess the statistical significance of differences observed between the pre- and post-intervention datasets.

Results

Students' knowledge of ethics in AI

Students' knowledge of ethics in AI was determined via their answers to the 12 multiple-choice questions from the pre- and post-questionnaires. Overall, the students demonstrated a moderate understanding of AI ethics ($M = 6.37$, SD $= 3.27$), which significantly improved following the educational intervention ($M = 10.26$, SD $= 1.57$), with statistical significance ($t(89) = 11.26$, $p < 0.001$, Cohen's $d = 1.19$).

To further analyze the data, an exploratory factor analysis with direct oblimin rotation was conducted, aiming to optimize the 12 knowledge questions into distinct factors. The decision to retain three factors was based on a combination of the Kaiser criterion (eigenvalues greater than 1), the scree plot, and the interpretability of the factor solution (Henson & Roberts, 2006). This analysis reduced the questions into three primary factors, which collectively account for 56% of the variation observed—considered reasonable in the humanities and social sciences (Goretzko et al., 2021). These factors were aligned with the initial categorization of the knowledge areas: risks associated with AI development and use (questions 1–4), potential solutions to these risks (questions 5–9), and broader ethical considerations in science and engineering domains (questions 10–12). Table 1 illustrates the t test results for each knowledge category, comparing scores from the pre-and post-intervention questionnaires, thereby illustrating the specific areas of knowledge improvement among the students.

Within the three knowledge domains assessed, the category 'Potential solutions to AI ethics concerns' registered the most pronounced improvement following the educational intervention. The mean scores in this

Table 1 Paired-sample *t* test results for students' knowledge of ethics in AI, pre- vs. post-results

Knowledge of ethics in AI (n = 90)	Pre-questionnaire		Post-questionnaire		*t*	*p*	*d*
	Mean	SD	Mean	SD			
Risks associated with AI development and use	1.74	1.60	3.23	0.89	8.52	< 0.001	0.90
Potential solutions to AI ethical concerns	2.56	1.43	4.56	0.67	12.85	< 0.001	1.35
Broader ethical considerations in science and engineering domains	2.07	0.96	2.47	0.66	3.59	< 0.001	0.38

category rose from 2.56 pre-intervention to 4.56 post-intervention, accompanied by the most substantial effect size reported (Cohen's *d* = 1.35). This significant increase indicates that the intervention was particularly effective in elevating students' understanding of how to address ethical issues in AI. Conversely, the smallest increase in knowledge was in the category 'broader ethical considerations in science and engineering fields'. It showed a modest increase, with mean scores ascending from 2.07 to 2.47, reflected by a relatively minor effect size of 0.38.

Students' perceived awareness of AI ethics

The analysis of students' responses to the eight statements representing perceived ethical awareness in the pre-and post-questionnaires indicated that prior to the intervention, they asserted a medium–high level of overall perceived awareness of ethical issues in AI ($M = 3.69$, SD = 0.56), which moderately increased after the intervention ($M = 4$, SD = 0.55). This difference was statistically significant with $t(89) = 6.60$, $p < 0.001$, Cohen's $d = 0.70$.

Students' self-reported ethical awareness, as measured through eight targeted statements in the pre- and post-intervention questionnaires, revealed an encouraging trend. Initially, the overall response indicated a medium–high awareness level ($M = 3.69$, SD = 0.56), which experienced a moderate yet significant increase following the

intervention ($M = 4$, SD = 0.55); this elevation was statistically significant with $t(89) = 6.60$, $p < 0.001$, and a robust effect size (Cohen's $d = 0.70$). Table 2 presents the pre- vs. post-questionnaire results for the eight statements representing students' perceived awareness of AI ethics.

The detailed comparison of the pre- and post-intervention responses illustrates an overall improvement in scores for all eight statements post-intervention. While all statements showed higher scores post-intervention, five statements showed statistically significant improvements, whereas the increases in three statements were not statistically significant the increases in three of these statements were not statistically significant (see Table 2). It is worth highlighting, however, that the last two statements, which probe students' confidence in identifying and addressing ethical dilemmas, initially garnered among the lowest scores. Interestingly, these two statements, despite their low initial scores, showed the highest increases from the pre-to the post-questionnaire, with statistical significance (Cohen's $d = 1.04$, 1.10, respectively).

Interestingly, the statement rated highest in both pre- and post-intervention questionnaires (3rd item) related to the importance of AI ethics education within academic programs, with similar and high scores in both the pre-questionnaire ($M = 4.43$, SD = 0.72) and the post-questionnaire ($M = 4.52$, SD = 0.60). The second highest

Table 2 Paired-sample *t* test results for students' perceived ethical awareness in AI

Perceived awareness (*n* = 90)	Pre-questionnaire		Post-questionnaire		*t*	*p*	*d*
	Mean	SD	Mean	SD			
1. Research should include aspects related to AI ethics	4.08	4.08	4.34	0.76	2.9	0	0.87
2. It is important to know more about ethical issues in AI	2.56	2.56	3.42	1	7.52	< 0.001	1.09
3. It is important to learn about ethical issues in AI during undergraduate studies	4.43	0.72	4.52	0.6	1.47	0.14	0.57
4. Ethics education may promote responsible behavior in AI	4.17	0.9	4.28	0.87	1.23	0.22	0.85
5. It is important to know more about ways to resolve ethical dilemmas in AI	4.3	0.94	4.5	0.69	2.16	0.03	0.88
6. I am familiar with ethical rules and regulations regarding AI development and use	4.37	0.77	4.44	0.75	1.07	0.29	0.69
7. I am confident in my ability to identify ethical dilemmas related to AI	3.02	1.07	3.43	0.91	3.76	< 0.001	1.04
8. I am confident in my ability to provide the best solution for ethical dilemmas related to AI	2.55	1.04	3.03	1.01	4.01	< 0.001	1.10

score pre-intervention (6th item) reflected students' perceived familiarity with AI ethical guidelines. Yet, it showed a slight, but not statistically significant, increase in the post-intervention scores.

Students' actual awareness of AI ethics

Prior to participation in the online explicit-reflective learning module, students demonstrated a moderate level of overall actual awareness of ethical issues in AI ($M = 2.68$, SD $= 0.96$), which significantly increased post-intervention ($M = 3.97$, SD $= 1.01$), with statistical significance and a high effect size $t(89) = 13.74$, $p < 0.001$, Cohen's $d = 1.45$.

A content analysis of pre-intervention responses yielded 151 text excerpts that identified ethical issues, while the post-intervention responses contained 277 text excerpts. This increase—from an average of 1.68 to 3.08 ethical issues identified per student—signifies that students were able to recognize a broader spectrum of ethical issues after completing the module.

The ethical issues were grouped into ten main categories, as detailed below:

1. *Risks to subjects*—Inadequate safeguards against risks to human participants, such as privacy breaches, physical and psychological harm, or adverse socioeconomic impacts.
2. *Biased data*—Utilization of inaccurate, incomplete, or incompatible databases, resulting in biases against certain individuals or societal groups.
3. *Security breach*—Insufficient protection against unauthorized data access or cyber-attacks, jeopardizing personal privacy.
4. *Manipulation*—Unauthorized reprocessing or sharing of user data beyond initial consent parameters.
5. *Disclosure ambiguity*—A lack of transparent and explicit communication to AI users about data collection, usage, and processing practices.
6. *Explainability*—Failure to interpret and explain the mechanisms by which AI systems turn inputs into outputs, thus the decisions that are made. This refers to both the creators and deployers of AI systems as well as to the end-users impacted by these systems.
7. *Lack of human autonomy*—Concerns that AI advancement may undermine human autonomy, compromising the capacity for independent decision-making and potentially conflicting with established ethical norms.
8. *Copyright issues*—Failure to provide explicit information on the ownership rights of data generated by AI systems, raising concerns about copyright infringement.
9. *Analysis errors*—Instances of human errors during data collection or analysis, mostly unintentionally.
10. *Accountability*—Absence of distinct mechanisms for assigning responsibility, which creates ambiguity regarding who bears accountability for the actions and decisions of AI systems.

Figure 1 depicts the distribution of the top five ethical issues in AI as identified by students, comparing responses from the pre- and post-questionnaires. This comparison shows a significant increase in the frequency of mentions across all categories, with statistical significance.
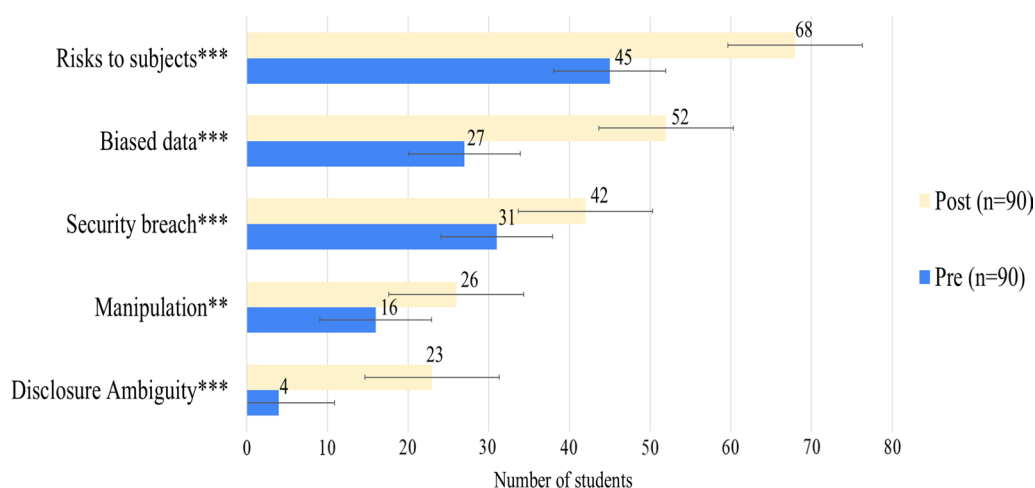

Fig. 1 Top-five categories for ethical issues indicated by students, pre vs. post-results. ***Significant at the 0.001 level (2-tailed); **Significant at the 0.01 level (2-tailed)

The most commonly identified category by students was 'risks to subjects'. One and a half times more students addressed this category in the post- vs. the pre-questionnaire (68 vs. 45 students). This difference was statistically significant with a large effect size ($Z = -4.80$, $p < 0.001$, $r = 0.51$). This increase indicates an enhanced student focus, from half to over three-quarters of the participants. Within this category, the predominant concern voiced by students centered around unauthorized access to personal information about users, highlighting fears of privacy breaches. As articulated by one student in the pre-questionnaire: "Big data may include private, personal, and possibly sensitive information of people, whose disclosure could harm their privacy and dignity, even if it is within the framework of scientific research" (S71, Aerospace Engineering).

Following the intervention, some students broadened their discussion on the 'risks to subjects', to include economic or psychological aspects. One student highlighted job displacement risks: "Depriving professionals of their jobs—people whose profession involves content writing are already in danger of losing their jobs to AI machines, and later also other professions such as customer service personnel, etc." (S14, Electrical Engineering). Another expressed concerns about involuntary personal revelations: "One particularly significant issue relates to the information that people will reveal about themselves, information that they may not want to know [..] it is not clear whether such information may harm those people and in what way this will affect them and their lives" (S43, Civil Engineering). These reflections demonstrate a deepened understanding and concern for the multifaceted ethical implications of AI on society.

The ethical issue of 'security breach' emerged as another significant concern, with its mentions increasing from 31 pre-intervention to 42 post-intervention, demonstrating statistical significance ($Z = -3.32$, $p < 0.001$, $r = 0.35$). Within this category, the majority of students expressed concerns regarding the vulnerability of data due to inadequate security measures. Students noted that the laboratory depicted in the picture appeared devoid of human presence, yet all the research materials were spread out across the room. This concern was articulated in responses such as: "Is the laboratory locked and the entrance to it secured in a way that ensures information security? The lab is empty of researchers, the laptop is open, maybe not locked with a password, and allows free access to data" (S47, Biomedical Engineering, pre-questionnaire).

This category was frequently associated with privacy concerns, with students emphasizing the importance of safeguarding against unauthorized access: "When it comes to a large database, it is important to make sure that there are protections against hacking attempts. This is the most important point in my view as it has serious consequences, especially when it comes to personal and sensitive information (which is often the case), the disclosure of which could seriously damage the right to privacy" (S85, Data Science, pre-questionnaire).

A third commonly mentioned category was 'biased data', with mentions nearly doubling post-intervention (52 post vs. 27 pre), with statistical significance and the highest effect size ($Z = -5$, $p < 0.001$, $r = 0.53$). Students voiced concerns over the dangers of relying on flawed or incomplete data sets, which could perpetuate discrimination against segments of the entire population. Concerns related to biased data were discussed in various sectors, including workplace discrimination, as exemplified in the following quote: "A decision-support algorithm used by workplaces may be biased according to gender, skin color, and other parameters that have nothing to do with the candidate's skills" (S61, Electrical Engineering, pre-questionnaire).

Additional concerns focused on the potential for discrimination in the judicial and law enforcement systems. For instance, one student expressed concern about facial recognition systems that might "identify people from a certain group in society as criminal suspects with a higher probability than the rest of the population" (S7, Electrical Engineering, pre-questionnaire). Another student pointed out potential biases in crime statistics: "If crime statistics show that a certain crime is more common among people of a certain origin, the interpretation of this could be racist and discriminatory" (S71, Aerospace Engineering, pre-questionnaire).

Participants underscored the role of discriminatory data in exacerbating social injustices. For example, one noted: "AI systems collect a lot of information from the network, the same network that perpetuates stereotypes" (S70, Data Science, pre-questionnaire). Another student expanded on this theme, suggesting that: "Decision-makers who will make use of information generated by AI will continue to perpetuate and preserve (even if not on purpose) discrimination and inequality in society" (S78, Medicine, post-questionnaire). Beyond discrimination, concerns about biased data influencing users' worldviews were raised. One student expressed this concern with respect to generative AI tools, stating: "The tool tends to adopt the values of OpenAI and produces an answer according to its (probably economic) worldview, which may lead to a very narrow and sometimes false perception of reality" (S37, Medicine, pre-questionnaire).

'Manipulation' was another frequently mentioned ethical issue. One and a half times more students addressed this category in the post- vs. the pre-questionnaire (26 vs. 16 students), with statistical significance ($Z = -3.16$,

$p = 0.00$, $r = 0.33$). Most students addressed the issue of manipulation concerning the transfer of user data to a third party for unethical purposes, such as financial, security, or political motives. Post-intervention responses reflected a tendency to interconnect various ethical issues, including information security failures and the manipulative use of data, for example: "There is a risk of leaking private information of the research participants. This information may be of value to political or commercial entities" (S15, Electrical Engineering, post-questionnaire).

The least-mentioned category among the top five ethical concerns was 'disclosure ambiguity', appearing in less than 5% of pre-questionnaire text excerpts. Yet, it showed a substantial increase in mentions during the post-questionnaire, rising to over a quarter of total mentions (23 mentions vs. 4 mentions, respectively), with statistical significance ($Z = -4.36$, $p < 0.001$, $r = 0.46$). Students stressed the importance of providing AI users with "a full and transparent explanation of how their data is collected and will be used in the future" (S57, Biomedical Engineering, post-questionnaire). Some expressed concern that users might be "unaware that they are in interaction with AI-based tools" (S78, Medicine, post-questionnaire).

Beyond the main categories mentioned above, five additional ones emerged in fewer responses, listed by descending frequency: explainability, human autonomy, copyright issues, analysis errors, and accountability. Each of these issues attracted 10% or fewer mentions.

Students' problem-solving skills related to AI ethics

Prior to participating in the online learning module, students demonstrated a relatively low proficiency in problem-solving skills related to AI ethics ($M = 1.99$, $SD = 0.71$). Post-intervention, there was an improvement in these skills ($M = 2.96$, $SD = 1$), demonstrating statistical significance and a high effect size ($t(89) = 11.09$, $p < 0.001$, Cohen's $d = 1.17$).

The content analysis of students' responses revealed a total of 91 text excerpts representing potential solutions to the ethical issues identified by students. This number nearly doubled, reaching 176 text excerpts in the post-questionnaire. This increase reflects a shift from an average of 1.01 to 1.96 suggested solutions per student, indicating an enhanced capacity to propose diverse remedies to ethical dilemmas post-module engagement.

The potential solutions were grouped into seven main categories as detailed below:

1. *Data security*—Adoption of advanced technologies and security measures to safeguard data against unauthorized access and cyber threats.

2. *Transparency*—Commitment to providing users with explicit and comprehensible information on how their data is collected, analyzed, and used.

3. *Technological quality control*—Application of automated techniques to mitigate algorithmic biases.

4. *Human quality control*—Utilization of human expertise for scrutinizing data collection and analysis methods, ensuring integrity and accuracy.

5. *Intra-organizational supervision*—Establishment of clear internal protocols and guidelines, including assigning ethics committees and conducting preliminary and periodic inspections, to uphold ethical standards.

6. *Intra-organizational employee training*—Equipping employees with the necessary knowledge and skills to navigate ethical considerations in their work, make informed decisions, and adhere to organizational ethical norms.

7. *External regulation*—Advocacy for and development of regulatory frameworks by public authorities. to promote and regulate the field of AI.

Figure 2 illustrates the distribution of the seven categories of proposed potential solutions to ethical issues in AI, as identified by students in the pre-and post-questionnaires. Notably, there was a substantial increase in mentions for all solution categories post-intervention, with statistical significance.

The most salient category emerging from both pre-and post-questionnaires was 'data security'. It was mentioned by 25 students before the participation and almost doubled with 42 mentions post-participation. This observed difference was statistically significant ($Z = -4.12$, $p < 0.001$, $r = 0.43$). Most students associated 'data security' as a pivotal response to the privacy-related issues outlined in the broader category of 'risks to subjects'. Recommendations centered on fortifying user privacy through enhanced laboratory and computer security measures, as articulated by one student: "Since there may be a substantial violation of the privacy of the people [..] there are several solutions for this: using a remote server with SSH and a password that allows access to the data only to the principal researcher [..]" (S9, Electrical Engineering, post-questionnaire). Several students also suggested "using various data anonymization techniques such as data masking, pseudonymization, and data swapping" as a recurrent solution to privacy concerns (S63, Mathematics, post-questionnaire).

'Intra-organizational supervision' stood out as the second most mentioned category, with its recognition doubling from the pre- to the post-questionnaire (18 to 35 students, respectively), with statistical significance ($Z = -4.12$, $p < 0.001$, $r = 0.43$). Students emphasized the
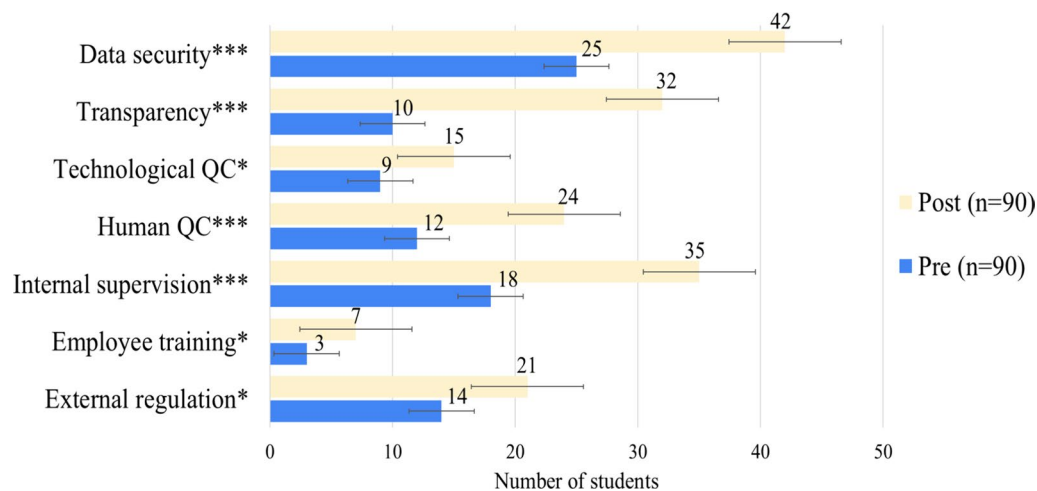
Fig. 2 All categories for possible solutions recognized by students, pre vs. post-results. ***Significant at the 0.001 level (2-tailed); **Significant at the 0.01 level (2-tailed); * Significant at the 0.05 level (2-tailed)

necessity for organizations engaged in AI research and development to establish "clear, transparent, and unambiguous internal protocols and procedures regarding how to engage and manage data" (S15, Electrical Engineering, post-questionnaire).

The 'transparency' category emerged as one of the most mentioned solutions, with threefold more mentions in the post-vs. the pre-questionnaire (32 vs. 10 students), achieving the highest effect size among the solution categories ($Z=-4.69$, $p<0.001$, $r=0.49$). Several students highlighted the importance of both the researcher and the researched signing an informed consent document. For example, one student emphasized: "Both sides should sign a document that provides complete transparency regarding the form of data collection, how the data will be used now and in the future, and whether there is an AI intervention" (S39, Mechanical Engineering, post-questionnaire). Others advocated for detailed transparency regarding AI system development and application processes: "To provide sincere answers regarding questions such as who took part in the development of the system, how was it implemented, what data was it trained on [..]" (S35, Biology, post-questionnaire).

The category of 'human quality control' emerged as the fourth most frequently mentioned solution in the post-intervention responses, doubling the mentions from 12 in the pre-questionnaire to 24 in the post-questionnaire, with statistical significance ($Z=-3.46$, $p<0.001$, $r=0.36$). Students advocated for human quality control as a versatile solution to a spectrum of ethical challenges, such as discrimination: "A possible solution depends on the researchers themselves [..] Employees should regularly review and re-evaluate AI

systems for biases and fairness, they should use diverse and representative data sets and employ techniques such as re-sampling, re-weighting, and data augmentation [..]" (S63, Mathematics, post-questionnaire).

Human quality control was additionally recognized as a key strategy to address explainability concerns. One student proposed the creation of "stopping points" in the research process to assess if an algorithm's explainability aligns with their expectations, highlighting a proactive approach to understanding AI decisions (S45, Data Sciences, pre-questionnaire). Moreover, one student identified this category as a potential solution to privacy and copyright issues, stating: "the researchers must be critical [..] make sure that it [the data] arrived in ways that do not contradict the requirement for privacy and do not harm the property of a particular person or organization" (S61, Electrical Engineering, post-questionnaire).

While 'external regulation' was among the less frequently mentioned solutions within the top five categories, it saw an increase in student responses from 14 in the pre-questionnaire to 21 in the post-questionnaire, demonstrating statistical significance ($Z=-2.64$, $p=0.01$, $r=0.28$). One student notably recommended: "I would suggest enacting laws that focus on the unique abilities of artificial intelligence [..].it is also necessary to create a regulatory supervision system in every factory or research laboratory that will make sure that the laws are followed" (S70, Data Science, pre-questionnaire).

In addition to the aforementioned five most common categories, two more areas, technological quality control' and 'employee training,' were identified in the

responses. However, these categories attracted relatively fewer mentions across both pre- and post-intervention questionnaires.

Discussion

The integration of AI ethics into higher educational curricula represents a central step towards preparing students to ethically navigate the complex landscape of AI technologies. This study embarked on a comprehensive examination of the integration of AI ethics within the educational curricula of science and engineering graduate students. The intervention, an online explicit-reflective learning module, was designed to enhance students' knowledge, ethical awareness, and problem-solving skills related to AI.

A noteworthy finding from this study is the substantial improvement in students' knowledge of AI ethics following the intervention, particularly in the comprehension of potential solutions to ethical challenges. This observation resonates with existing literature indicating the positive effects of AI ethics education on knowledge enhancement (Kong et al., 2023; Ouyang et al., 2022). This enhancement could suggest that participation in the online explicit-reflective learning module deepened students' comprehension of ethical issue navigation within AI context, an encouraging sign that such educational programs can have a pronounced effect on practical aspects of ethics in AI.

Regarding perceived ethical awareness, initially, students self-reported a medium–high level of awareness concerning AI ethical issues, which saw a statistically significant moderate increase post-intervention. Notably, the statements that gauged student confidence in identifying and addressing ethical issues initially garnered among the lowest scores. However, these areas witnessed the most marked improvement following the intervention. Moreover, the statement rated highest in both pre- and post-intervention questionnaires emphasized the critical importance of integrating AI ethics education within academic programs. This consistent recognition underscores a broad consensus among students about the essential role of ethics education in preparing them for the ethical challenges of AI technologies. Such acknowledgment implies that students are not only aware of but also value the need for a deep understanding of AI ethics as an integral part of their academic and professional development.

The identification of the types of ethical issues and possible solutions identified by students provides valuable insights into their actual ethical awareness and problem-solving skills. Through content analysis of open-ended responses, ten key ethical issues and seven key solution categories emerged. Although no new categories emerged in the post-questionnaire, each of the existing categories showed an increase in the number of mentions, with statistical significance. This suggests that the course module reinforced and deepened students' understanding of the key ethical issues and possible solutions. In addition, the examples provided by students' post-intervention were more detailed and focused, reflecting a more nuanced understanding of the ethical implications.

Before exposure to the module, students already demonstrated a moderate level of actual awareness, identifying ten distinct categories of ethical issues in AI. For example, the category of 'disclosure ambiguity' received only four mentions in the pre-questionnaire. This category was highlighted in the case studies presented in the online module. Consequently, this category witnessed five times more mentions in the post-questionnaire, indicating that more students were exposed to this ethical issue through the course content.

The category of 'risks to subjects' was the most cited ethical issue, highlighting concerns about privacy and unauthorized access to personal information. This underscores students' growing awareness of AI's ethical implications on society. Another prominent concern was 'security breach', where students noted data vulnerability due to insufficient security measures. These concerns align with prior research identifying privacy and data protection as critical ethical issues in AI development (Holmes et al., 2022; Jobin et al., 2019). For addressing these concerns, 'data security' emerged as the primary solution suggested by students, focusing on protecting sensitive information through methods like data encryption. This approach is consistent with strategies recommended in the literature for preserving privacy by anonymizing data sets (Rocher et al., 2019).

Further ethical issues mentioned by students included 'biased data' and 'manipulation', which saw a significant increase in mentions post-intervention. Students expressed worries about the potential for discrimination based on flawed data sets and unethical transfer of user data to third parties for financial, security, or political purposes, mirroring previous findings (Borenstein & Howard, 2021; Pavlik, 2023).

Lastly, 'disclosure ambiguity', highlighting the need to provide AI users with clearer explanations of data collection and usage, was among the least cited ethical issues. This is in contrast to prior studies that reported on transparency and accountability as the most prominent ethical concerns in AI (Holmes et al., 2022; Zhou et al., 2020). Yet, this category showed a substantial increase in mentions during the post-questionnaire. This discrepancy underscores the evolving landscape of AI ethics education and the need for continued emphasis on transparent

AI development and user consent processes (Akgun & Greenhow, 2022; Bogina et al., 2022).

Another prominent finding relates to student problem-solving skills in AI ethics. Prior to the participation, students demonstrated a relatively low proficiency in problem-solving skills compared to their actual awareness in terms of identifying ethical issues. Yet, this proficiency experienced an enhancement following their participation.

Impressively, every identified category of potential solutions experienced a notable increase in mentions following the intervention. The content analysis revealed that 'data security' surfaced as a predominant concern, underscoring its essential role in addressing privacy risks—a core component within the broader ethical consideration of 'risks to subjects'. This concern reflects a deeper student understanding of the necessity for robust security measures to protect user privacy, aligning with contemporary scholarly emphasis on data security in AI ethics discussions, as articulated in prior studies (Akgun & Greenhow, 2022; Borenstein & Howard, 2021; Jobin et al., 2019).

Similarly, the importance of 'intra-organizational supervision' was significantly enhanced, with students recognizing the imperative for organizations to develop clear, transparent protocols for data management, echoing the literature's call for transparent organizational practices in AI research and development (Erduran, 2023; Holmes et al., 2022; Taddeo & Floridi, 2018). It seems that while suggesting possible solutions to AI issues, students focus predominantly on immediate, actionable strategies that can be implemented within organizational and legal frameworks.

Overall, the study's results indicate a significant shift in students' ethical awareness and problem-solving skills following the intervention, indicating an enhanced ability to address and respond to the ethical challenges associated with AI technologies effectively. Such a synthesis not only highlights the need for thorough ethics instruction within AI education frameworks but also enriches the ongoing academic discourse on cultivating a generation of professionals who are well-equipped to navigate the ethical landscapes of AI with informed judgment and integrity.

Based on these findings, we aim to further develop the AI-ethics module to enhance its effectiveness. Specifically, we plan to update the module to align with new emerging technologies and present new case studies that reflect current ambiguous situations, such as facial recognition technologies for campus security and personalized chatbots for education. Furthermore, we will enhance the interactive elements of the module to promote deeper engagement and critical thinking by incorporating activities such as role-playing scenarios and interactive simulations. Lastly, feedback from students who took the online module in its current form will also be used to continuously refine and improve the module.

While this research offers significant insights into the integration of AI ethics in educational science and engineering curricula, it is accompanied by limitations that warrant consideration. Conducted within a particular context—graduate students at a leading technological university—the study's applicability to other disciplines or educational levels may be limited. In addition, the study did not include a comparison or experimental group, which limits the ability to make causal inferences. To broaden our understanding of AI ethical awareness, further research in varied educational settings and across diverse student demographics is essential. Moreover, since the effect of the intervention was assessed at the end of the course, it is recommended to investigate the long-term impact by examining the effects over an extended period. It is recommended that future studies expand across different contexts and disciplines as well as include longitudinal assessments to determine the enduring influence of AI ethics education.

Conclusions and implications

Our work aims to expose students to the ethical and societal implications of AI, thereby arming future scientists and engineers with the ethical frameworks necessary for responsible innovation in the digital age. Furthermore, our findings provide a blueprint for establishing a supportive organizational environment conducive to the advancement of AI ethics education within academic institutions.

The study's findings offer insights for educators and curriculum designers aiming to integrate AI ethics into higher education curricula. It underscores the role of an online module on AI ethics that blends case-based learning with reflective exercises in fostering students' ethical knowledge, awareness, and problem-solving skills. Implementing similar educational strategies, including workshops, seminars, and online modules can profoundly impact students' readiness to address the real-world challenges they will probably face in their careers.

Given the documented improvement in students' ethical awareness and problem-solving skills following the intervention, educators and curriculum designers should advocate for the integration of AI ethics into compulsory curricula for science and engineering disciplines. This would ground a baseline level of ethical competence among future professionals. Furthermore, fostering collaborations between educational institutions, the tech industry, and governmental bodies can amplify the practical relevance of AI ethics education. By co-developing

case studies and educational materials with tech companies, students can gain direct insight into the ethical dilemmas currently facing the sector, preparing them for responsible AI use that considers both individual and societal stakes.

Availability of data and materials
The data sets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Ethical approval was received by the local Ethics Committee of the Technion—Israel Institute of Technology. All students gave their consent to participate in the study and signed a consent form.

Consent for publication
The manuscript does not include images or videos relating to an individual person. All the figures (graphs) that are included in the manuscript were generated by the authors and the journal has our consent to publish them.

Competing interests
The authors have no competing interests to declare that are relevant to the content of this article.

References

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI Ethics, 2*, 431–440. https://doi.org/10.1007/s43681-021-00096-7

Bairaktarova, D., & Woodcock, A. (2017). Engineering student's ethical awareness and behavior: A new motivational model. *Science and Engineering Ethics, 23*, 1129–1157. https://doi.org/10.1007/s11948-016-9814-x

Barak, M., & Green, G. (2020). Novice researchers' views about online ethics education and the instructional design components that may foster ethical practice. *Science and Engineering Ethics, 26*(3), 1403–1421. https://doi.org/10.1007/s11948-019-00169-1

Barak, M., & Green, G. (2021). Applying a social constructivist approach to an online course on ethics of research. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-021-00280-2

Barak, M., & Usher, M. (2019). The innovation profile of nanotechnology team projects among face-to-face and online learners. *Computers & Education, 137*, 1–11. https://doi.org/10.1016/j.compedu.2019.03.012

Bogina, V., Hartman, A., Kuflik, T., & Shulner-Tal, A. (2022). Educating software and AI stakeholders about algorithmic fairness, accountability, transparency and ethics. *International Journal of Artificial Intelligence Education, 32*, 808–833. https://doi.org/10.1007/s40593-021-00248-0

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics, 1*(1), 61–65. https://doi.org/10.1007/s43681-020-00002-7

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: a systematic literature review. *International Journal of STEM Education*. https://doi.org/10.1186/s40594-023-00418-7

Cooper, G. (2023). Examining science education in ChatGPT: an exploratory study of generative artificial intelligence. *Journal of Science Education and Technology, 32*, 444–452. https://doi.org/10.1007/s10956-023-10039-y

Creswell, J. W. (2014). *A concise introduction to mixed methods research*. Sage Publications.

Erduran, S. (2023). AI is transforming how science is done. Science education must reflect this change. *Science*. https://doi.org/10.1126/science.adm9788

Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology, 40*, 3510–3521. https://doi.org/10.1007/s12144-019-00300-2

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*, 393–416. https://doi.org/10.1177/0013164405282485

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., et al. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence Education, 32*, 504–526. https://doi.org/10.1007/s40593-021-00239-1

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems*, First Edition. IEEE, 2019. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kong, S.-C., Cheung, W.M.-Y., & Zhang, G. (2023). Evaluating an artificial intelligence literacy programme for developing university students' conceptual understanding, literacy, empowerment and ethical awareness. *Educational Technology & Society, 26*(1), 16–30.

Kumar, R., Eaton, S. E., Mindzak, M., & Morrison, R. (2024). Academic integrity and artificial intelligence: An overview. In S. E. Eaton (Ed.), *Second handbook of academic integrity*. Springer International Handbooks of Education. https://doi.org/10.1007/978-3-031-54144-5_153

Lin, X.-F., Wang, Z., Zhou, W., Luo, G., Hwang, G.-J., Zhou, Y., et al. (2023). Technological support to foster students' artificial intelligence ethics: An augmented reality-based contextualized dilemma discussion approach. *Computers & Education*. https://doi.org/10.1016/j.compedu.2023.104813

Martin, D. A., Conlon, E., & Bowe, B. (2021). Using case studies in engineering ethics education: The case for immersive scenarios through stakeholder engagement and real life data. *Australasian Journal of Engineering Education, 26*(1), 47–63. https://doi.org/10.1080/22054952.2021.1914297

Mitcham, C., & Englehardt, E. (2019). Ethics across the curriculum: Prospects for broader (and deeper) teaching and learning in research and engineering ethics. *Science and Engineering Ethics, 25*, 1735–1762. https://doi.org/10.1007/s11948-016-9797-7

Mouta, A., Torrecilla-Sánchez, E. M., & Pinto-Llorente, A. M. (2019). Design of a future scenarios toolkit for an ethical implementation of artificial intelligence in education. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12229-y

Nam, B. H., & Bai, Q. (2023). ChatGPT and its ethical implications for STEM research and higher education: a media discourse analysis. *International Journal of STEM Education*. https://doi.org/10.1186/s40594-023-00452-5

OECD Council, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449 5 (Adopted on May 22, 2019; 2021), available at https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies, 27*, 7893–7925.

Park, J., Teo, T.W., Teo, A., Chang, J., Huang, J.S., & Koo, S. (2023). Integrating artificial intelligence into science lessons: teachers' experiences and views. *International Journal of STEM Education*. https://doi.org/10.1186/s40594-023-00454-3

Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator*. https://doi.org/10.1177/10776958221149577

Qiao, C., Chen, Y., Guo, Q., & Yu, Y. (2024). Understanding science data literacy: A conceptual framework and assessment tool for college students majoring in STEM. *International Journal of STEM Education*. https://doi.org/10.1186/s40594-024-00484-5

Raji, I. D., Scheuerman, M. K., & Amironesei, R. (2021). You can't sit with us: Exclusionary pedagogy in AI ethics education. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 515–525. https://doi.org/10.1145/3442188.3445914

Rocher, L., Hendrickx, J. M., & de Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communication, 10*, 3069. https://doi.org/10.1038/s41467-019-10933-3

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science, 361*(6404), 751–752. https://doi.org/10.1126/science.aat5991

Taebi, B., & Kastenberg, W. E. (2019). Teaching engineering ethics to PhD students: A Berkeley-Delft initiative. *Science and Engineering Ethics, 25*, 1763–1770.

Usher, M., & Barak, M. (2020). Team diversity as a predictor of innovation in team projects of face-to-face and online learners. *Computers & Education*. https://doi.org/10.1016/j.compedu.2019.103702

Usher, M., & Hershkovitz, A. (2022). Interest in educational data and barriers to data use among Massive Open Online Course instructors. *Journal of Science Education and Technology, 31*, 649–659. https://doi.org/10.1007/s10956-022-09984-x

Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: a systematic review from 2011 to 2021. *International Journal of STEM Education*. https://doi.org/10.1186/s40594-022-00377-5

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*. https://doi.org/10.1186/s41239-019-0171-0

Zhou, J., Chen, F., Berry, A., Reed, M., Zhang, S., & Savage, S. (2020). A survey on ethical principles of AI and implementations. *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. Canberra, ACT, Australia, https://doi.org/10.1109/SSCI47803.2020.9308437

Publisher's Note