Teaching postsecondary students about the ethics of artificial intelligence: A scoping review protocol

Calvin Hillis[1], Maushumi Bhattacharjee[2], Batool AlMousawi[3], Tarik Eltanahy[1], Sara Ono[1], Marcus Hui[4], Ba' Pham[5], Michelle Swab[6], Gordon V. Cormack[7], Maura R. Grossman[7], Ebrahim Bagheri[8¶*], Zack Marshall[9,10¶]

[1] The Creative School, Toronto Metropolitan University, Toronto, Ontario, Canada

[2] Faculty of Law, McGill University, Montreal, Quebec, Canada

[3] Dalla Lana School of Public Health, University of Toronto, Ontario, Canada

[4] Department of Biology, Queens University, Kingston, Ontario, Canada

[5] Toronto Health Economics and Technology Assessment Collaborative, University of Toronto, Toronto, Ontario, Canada

[6] Health Sciences Library, Memorial University, St. John's, Newfoundland and Labrador, Canada

[7] David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada

[8] Faculty of Information, University of Toronto, Toronto, Ontario, Canada

[9] Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada

[10] School of Social Work, McGill University, Montreal, Quebec, Canada

* Corresponding author

E-mail: ebrahim.bagheri@utoronto.ca (EB)

¶ EB and ZM are Joint Senior Authors.

Abstract

The field of AI carries inherent risks such as algorithmic biases, security vulnerabilities, and ethical concerns related to privacy and data protection. Despite these risks, AI holds significant promise for social good, with applications ranging from improved healthcare diagnostics to enhanced education strategies. Teaching AI ethics in postsecondary settings has emerged as one of the strategies to mitigate AI-related harms. The objectives of this review are to (1) synthesize existing research related to teaching postsecondary students about the principles and practice of ethics and AI, and (2) identify how educators are evaluating changes in student knowledge, skills, attitudes, and behaviors. This scoping review will follow the first five steps articulated by Arksey and O'Malley. A structured search strategy developed by an academic librarian incorporates three primary concept groups related to education, AI, and ethics. Database search strategies emphasize sensitivity rather than precision, given that a supervised machine learning tool will be used to assist in the identification of relevant abstracts. Searches will be conducted in the following academic databases: PubMed, Embase, Scopus, ERIC, LISTA, IEEE Xplore, APA PsycInfo, and ProQuest Dissertations and Theses. Results will include an up-to-date synthesis of the current state of AI ethics education in postsecondary curricula, evaluated teaching strategies, and potential outcomes associated with AI ethics education. Search results will be reported according to the PRISMA-ScR checklist. Data charting will focus on AI ethics pedagogy. This review will inform future research, policy development, and teaching practices, offering valuable insights for educators, policymakers, and researchers working towards responsible AI integration. Findings will contribute to enhanced understandings of the complexities of AI ethics education and have the potential to shape the ways trainees in multiple disciplines learn about the ethical dimensions of AI in practice.

Introduction

The field of Artificial Intelligence (AI) is booming. While the term "artificial intelligence" was first coined at the 1956 Dartmouth conference [1], the launch of ChatGPT in November 2022 catalyzed newfound interest in AI including changes in public opinion [2]. Academic and public discourse highlights the potential impacts of AI, including societal transformation, huge potential benefits, and imminent risks [3]. In Roe and Perkins' [4 p1] recent discourse analysis of AI in UK news media headlines, "…results show that there is a complex and at times paradoxical portrayal of AI in general and ChatGPT as well as other Large Language Models". Most recently, the Group of Seven (G7) leaders introduced the Hiroshima AI Process including International Guiding Principles and a Code of Conduct for organizations developing advanced AI systems [5]. As AI gains increasing prominence in the public consciousness, there is a growing imperative for the establishment of policies to ensure that the applications of AI systems are firmly rooted in ethical considerations.

The field of AI carries inherent risks, encompassing technical challenges such as algorithmic biases, security vulnerabilities, and ethical concerns related to privacy and data protection [6,7]. Despite these risks, AI holds significant promise for social good, with applications ranging from improved healthcare diagnostics to enhanced educational strategies [8-10]. However, ethical considerations are paramount to mitigate potential harm [11], requiring transparent and accountable algorithmic decision-making, proactive efforts to address biases in training data, and robust security measures [12]. Striking a delicate balance between realizing the transformative benefits of AI and managing its associated risks is crucial for a responsible and impactful integration of AI technologies into society.

AI ethics education is one of the proposed approaches to mitigating AI-related harms

[11]. By integrating ethics into AI education, future AI practitioners across disciplines will be able to develop fair, accountable, and transparent AI systems [11]. The proposed scoping review is based on the following assumptions: (1) AI ethics content is integrated in relevant postsecondary training including computer science, engineering, business, health, social sciences, and humanities curricula, (2) strategies for teaching about AI ethics have been evaluated, and (3) teaching about AI ethics leads to better outcomes. We will explore each of these assumptions in turn.

AI and ethics in course content

The teaching of AI ethics varies across disciplines and is informed by different paradigmatic assumptions and ideologies. This lack of standardization hinders interdisciplinary collaboration and consistency in curricula. For example, ethical philosophy and its jargon is used in Humanistic Social Science (HSS) courses as assumed knowledge. Computer Science (CS) courses often do not translate technical details, programming, and equations to be discernible to non-majors [12]. In addition, CS courses often reference industry-specific guidelines such as the Association for Computing Machinery (ACM) Code of Ethics, while HSS content focuses on the human impact of AI systems [12]. Thus, AI ethics discourse often adopts discipline-specific norms and philosophical assumptions. The “constructed distinctions” surrounding each discipline prompt inherent separation [12].

In a review of 94 international CS ethics course syllabi, researchers reflected on the ethical considerations taught to students [13]. They found that courses generally consisted of the instructor’s relativist conception of “doing good”, the traditions of CS, and the tensions that arise while navigating industry demands and societal concerns [13]. This relativist influence was also

noted in a syllabi review of Australian universities' CS ethics courses [14]. The review found no standardized macro-ethical agenda to guide AI ethics training and evaluation. Similarly in an international review, Suárez and Varona [15] analyzed 503 course syllabi across 66 universities and 16 countries. Their results align with previous findings, underlining how existing AI ethics curricula do not conform to a universal set of circulatory norms [15]. Of the 503 course syllabi examined, 84.69% had poorly articulated learning outcomes, little to know overlap, and minimal differentiation between the knowledge, values, and goals to be taught. Overall, research suggests that the existing approaches to teaching ethics of machine learning and AI lack standardized structure and cohesive goals, with no universal definition of "ethics" [12,16,17]. It is important to note that this variation is not necessarily counter-productive to AI ethics discourse but rather reflects the diverse value systems from around the world that AI ethics discourse can draw from to better serve diverse populations.

Strategies for teaching about ethics and AI

Javed and colleagues [16] used a method called Latent Dirichlet Allocation (LDA) to analyze how AI ethics is taught in university courses. They looked at 166 courses from different parts of the world, focusing on who teaches them, where they are taught, and at what cognitive level [16]. Their analysis identified the ways AI ethics is taught from different technical, legal, and value-based perspectives depending on field of study [16]. In a separate study about strategies for teaching AI ethics, Raji, Scheuerman, and Amironesei [12] observed a rigid adherence to specific methodologies, influenced by disciplinary preferences, contributing to a separation between the HSS, which tends to favor qualitative methods, and STEM disciplines that value quantitative and computational approaches. The lack of efforts to bridge this gap

seems to worsen the division. In HSS courses, there is an assumption of prior knowledge of ethical philosophy, while CS courses struggle to present technical details in a way that is accessible to non-majors. This divergence in the discourse on AI ethics, shaped by industry-specific guidelines in CS and a focus on human impact in HSS, influences and is reflected in approaches to teaching about ethics and AI.

Pedagogical strategies employed in CS ethics education may point to effective approaches to teaching AI ethics. Simulations and role-playing have been shown to be effective in CS ethics instruction [17,18]. Simulations reflecting real-life situations enhance students' ethical capacities and interest in CS ethics through technical courses [11,13,17,18]. Role-playing activities prompt students to translate ethical decisions into code, assess biases, and experience decision-making from various perspectives in algorithmic system development, emphasizing the importance of diverse considerations [18]. The integration of role-playing into "in situ" approaches, exemplified by courses like "Deep-tech ethics" [19], significantly enhances students' conceptual understanding and engagement with ethics. There are instances of these pedagogical strategies being applied in AI ethics contexts, though a scoping review that captures these educational trends has yet to be carried out.

Outcomes of teaching about ethics and AI

This review aims to consolidate existing knowledge on teaching postsecondary students the principles and application of ethics within the development and utilization of AI. While some studies have presented learning outcomes and discussed assessment methods, there is a noticeable lack of comprehensive information on this topic [20,21]. Scholars have shared insights on the desired learning outcomes for university students in AI ethics courses. These

encompass heightened awareness of ethical considerations in AI development, fostering collaborative understanding among students from diverse backgrounds, refining critical assessment skills for ethically informed decisions in AI, applying ethical theories to AI-related dilemmas, recognizing algorithmic biases, incorporating ethical considerations into AI system design, showcasing ongoing interest in AI ethics beyond the course, and more [11,22-24]. Evaluating these outcomes may involve a blend of quantitative measures (e.g., surveys, exams) and qualitative assessments (e.g., class discussions, projects) [18,24]. Incorporating real-world case studies can offer a practical evaluation of students' proficiency in applying AI ethics principles in intricate situations [8,24,25]. Some scholars critique the impact of AI ethics principles, noting their often-ineffectual nature in practice, failing to address the societal damages of AI technologies [26]. This gap between principles and technological reality poses risks, diverting resources away from potentially more impactful endeavors [26]. To address these challenges, Munn suggests exploring alternative approaches to AI justice that extend beyond ethical principles, encompassing a broader consideration of oppressive systems and a focused examination of accuracy and auditing [26].

Materials and methods

Despite the volume of research in this field, this scoping review is the first to document what is known about teaching postsecondary students about the ethics of AI, with particular attention to course content, teaching strategies, identified learning outcomes, and approaches to evaluating student learning. Drawing on qualitative, quantitative, and mixed methods research, this review will be conducted in accordance with the scoping review methods articulated by Arksey and O'Malley [27] including the first five steps: (1) identifying the research questions;

(2) identifying relevant studies; (3) study selection; (4) charting the data; and (5) collating, summarizing, and reporting the results. A Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) checklist has been completed (Supporting Information) [28]. A Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist [29] will be submitted with scoping review results.

Step 1: Identifying the research questions

The objectives of this review are to: (1). Synthesize existing research related to teaching postsecondary students about principles and practice of ethics in the development and use of AI, with particular attention to fairness, accountability, transparency, bias, and explainability. (2) Identify how educators are evaluating changes in student knowledge, skills, attitudes, and behaviors consistent with evolving social expectations and values towards AI.

Two research questions have been identified:

(1) What is known about teaching postsecondary students about the principles and practice of ethics in the development and use of AI?

(2) How can students gain the knowledge, skills, attitudes, and behaviors consistent with evolving social expectations and values towards AI?

Key concepts

This review focuses on teaching ethics to postsecondary students for the ethical use and development of AI. Each concept is defined below.

Artificial Intelligence. The term “artificial intelligence” has evolved tremendously since its conception, moving from Arthur Lee Samuel’s simple checkers program to complex

transformer neural networks such as Open AI's Chat GPT [4]. For purposes of this review, AI includes reference to artificial intelligence, computing, deep learning, computing algorithms, machine learning, natural language processing, and language learning.

Ethics. The concept of AI ethics encompasses the ethical considerations and principles guiding the development, deployment, and use of AI systems. Several guidelines have been proposed to address AI ethics [30], with attention to key dimensions such as accountability, algorithmic bias, fairness, transparency, and explainability [31]. Each concept is defined below.

Accountability in AI ethics refers to the responsibility and answerability of individuals, organizations, and systems for the impact of AI technologies on society [32]. It involves defining clear lines of responsibility for the development, deployment, and consequences of AI systems. Frameworks such as the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems [33] emphasize the importance of accountability in AI decision-making processes.

Algorithmic bias refers to the unintentional and discriminatory outcomes produced by algorithms and data-driven decision-making systems [34]. This bias may arise from historical discrimination embedded in data sources, resulting in the inadvertent amplification of existing biases. Even when sensitive attributes are omitted, correlations within the data can lead to biased algorithmic behavior through proxy variables. Addressing algorithmic bias requires a proactive and discrimination-conscious approach in the design and implementation of data mining systems.

Algorithmic fairness encompasses metrics, methods, and representations aimed at ensuring equitable and unbiased outcomes in algorithmic decision-making [34,35]. This field emerges from the growing influence of algorithms in diverse domains and considers the socio-technical impacts of bias and discrimination [36]. Algorithmic fairness extends beyond traditional technical domains, integrating insights from social sciences, law, economics,

philosophy, sociology, political science, and communication. The goal of algorithmic fairness is to create decision-making systems that consider ethical and societal implications, reducing bias and increasing equity in their outputs.

Transparency in AI ethics pertains to the openness and comprehensibility of AI systems. Transparent AI systems enable users, stakeholders, and the public to understand how decisions are made [37]. Explainability involves the ability to understand and interpret how AI systems arrive at specific decisions or predictions [31,37]. Explainable AI (XAI) serves as a pivotal subset of AI, aiming to demystify the functioning of complex models and algorithms [37]. The primary objective of XAI is to bring transparency to the decision-making processes of various machine learning algorithms [37].

Postsecondary Education. For the context of this review, postsecondary education includes college and university settings, including diploma, undergraduate, and graduate training.

Step 2: Identifying relevant studies

The structured search strategy for this review was developed by an academic librarian in consultation with the review team. Based on the concepts identified above, the search incorporates three primary concept groups that include both keyword terms applied to database fields such as titles and abstracts, as well as controlled vocabulary terms where applicable. The first concept group consists of words relating to education, such as “teaching”, “universities”, and “postsecondary”. The second concept group includes words relating to ethics, and the third concept group relates to AI. While there are many synonyms and related words for AI, terms such as “machine learning” or “artificial intelligence” are typically used in the article description or abstract. The three concept groups were combined using Boolean AND. Database search

strategies emphasize sensitivity rather than precision [38], given that a supervised machine learning tool [39] will be used to assist in the identification of relevant abstracts.

An initial limited search of PubMed and Scopus was undertaken to identify articles on the topic. The text words contained in the titles and abstracts of relevant articles, and the index terms used to describe the articles were used in the development of a sample search strategy for PubMed (S1 Table). The initial search strategy, including all identified keywords and index terms, has been adapted for each included database. No language limiters will be imposed during the database searches. In addition, publication date limiters will not be used as the discussion of ethics in AI has been ongoing for several decades.

Sources to be searched include: PubMed (Medline), Embase (via embase.com), Scopus, ERIC (via EBSCOhost), LISTA (via EBSCOhost), IEEE Xplore, APA PsycInfo (via EBSCOhost), and ProQuest Dissertations and Theses.

Step 3: Study selection

All identified citations will be collated and uploaded into the Continuous Active Learning® tool (CAL®), which uses supervised machine learning to rank titles and abstracts based on relevance to the search [39]. CAL® is a machine-learning tool that continuously re-ranks the abstracts based on screening decisions, so that the abstracts that are most likely to meet the inclusion criteria are presented before those that are less likely [39]. Screening ends when the remaining abstracts are very unlikely to meet the inclusion criteria. CAL® has been demonstrated to have comparable or superior accuracy to that which would have been achieved by exhaustive manual screening of every retrieved abstract, with much less human effort [39]. Because CAL® presents for screening only the abstracts that are reasonably likely to meet the

criteria, it is possible to employ broad searches that retrieve more abstracts than could reasonably be screened, and there is less need to deduplicate the search results from multiple databases prior to screening.

After a pilot testing phase, each reference will be screened by a team of two independent reviewers for assessment against the review inclusion criteria. Reviewers will manually keep track of their decisions. Once they have both screened five consecutive sets of 100 references in which the inclusion rate is below 5%, screening on title and abstract will stop, as recommended by the CAL® developers. Screening results will then be extracted from the CAL® system and imported into EPPI-Reviewer [40]. This will include four separate RIS files: Reviewer (1) Includes, Reviewer (1) Excludes, Reviewer (2) Includes, Reviewer (2) Excludes. Within EPPI-Reviewer, using a Boolean search, any differences between Reviewer (1) and (2) will be identified and reviewed by a third member of the team.

Study selection criteria: screening on title and abstract

The inclusion criteria for screening on title and abstract are:

(1) Language: Written in English. Due to team capacity only references published in English will be included in the review. Any non-English references will be identified in a separate category and the number of references in this group will be clearly reported.

(2) Setting: Postsecondary education (college or university)

(3) Topic: AI AND ethics

The exclusion criteria for screening on title and abstract are:

(1) Language: The article is not written in English

(2) Setting: Primary or secondary school (K-12), or non-educational setting such as industry

(3) Topic: Not related to AI OR not related to ethics

Study selection criteria: screening on full text

Sources included after screening on title and abstract will be retrieved and screened on full text in EPPI-Reviewer. The full article will be assessed in detail against the inclusion criteria by two or more independent reviewers. Reasons for exclusion of sources of evidence at full text will be recorded and reported in the scoping review. Any disagreements that arise between the reviewers at each stage of the screening process will be resolved through discussion, or with an additional reviewer. The results of the search and the screening process will be reported in full in the final scoping review and presented using a PRISMA flow diagram.

The inclusion criteria for screening on full text are:

(1) Language: Written in English

(2) Setting: Postsecondary education (college or university)

(3) Topic: Ethical AI pedagogy/curriculum

(4) Focus: Ethical AI pedagogy/curriculum is the primary focus of the paper

The exclusion criteria for screening on full text are:

(1) Language: The article is not written in English

(2) Setting: Primary or secondary school (K-12), or non-educational setting such as industry

(3) Topic: Not related to AI OR not related to ethics

(4) Format: Conference / tutorial workshop (outside of postsecondary environment)

(5) Format: AI ethics education review

(6) Format: Rationale / argument for teaching about ethical AI (why this is needed)

(7) Focus: Student diversity/inclusion in AI education

(8) Focus: Student perspectives on ethical AI

While important, articles that are excluded due to Format do not provide the specific topic insight that we are seeking in this review. An example of a reference coded as "AI Ethics education review" is "More than 'if time allows': the role of ethics in AI education" [20]. This article analyzes a selection of AI ethics and technical AI courses to understand which ethics related topics instructors cover in their courses. Similarly, "The need for health AI ethics in medical school education" [41] would be coded as "Rationale / arguments for teaching about ethical AI". This article notes a technological shift in the medical field, highlights the potential of AI in healthcare, and provides an argument for introducing AI ethics instruction in medical school. While useful for the discipline, papers such as these typically review AI course syllabi, or emphasize the importance of teaching AI ethics, while our review is primarily focused on collecting data from the peer-reviewed literature about pedagogical approaches to AI ethics instruction.

Step 4: Charting the data

Data charting will be done manually by one team member and verified independently by a second team member. If the second reviewer has questions or concerns, the two team members will work together to determine eligibility and inclusion for data synthesis and charting.

Data-charting elements

The following information will be charted from each included study: (1) article format; (2) curriculum/course content; (3) educational setting; (4) educational strategies; (5) learning outcomes evaluation method; (6) learning outcomes analysis method; (7) reported learning

outcomes; and (8) outcomes direction of change.

Article Format. Reviewers will indicate relevant elements related to course format including: (1) detailed course plan provided; (2) general course information provided; (3) conference tutorial/workshop outside of postsecondary environment; (4) rationale/arguments for teaching about ethical AI; (5) argument for teaching AI ethics in a particular way; (6) integrated ethical AI content (ethical AI content integrated with technical teaching); (7) stand-alone ethical AI content (course specifically about ethics); (8) content specific ethics (e.g., data ethics); and (9) description of course activity.

Course / Curriculum Content. Reviewers will select from the following options: (1) fairness, accountability, and transparency (FAccT) combined concept; (2) fairness; (3) accountability; (4) transparency; (5) algorithmic bias; (6) privacy; (7) equity; (8) trust; (9) social responsibility; (10) general ethics/ethical thinking/ethical concepts; (11) safety; (12) security; (13) AI harms; (14) governance or regulation; (15) bias; (16) explainability; and (17) other.

Educational Setting. Reviewers will choose from the following educational settings: (1) undergraduate setting; (2) graduate setting; (3) computer science/engineering in general; (4) non-computer science; (5) other; and (6) not specified.

Educational Strategies. Possible educational strategies include: (1) case study; (2) real-world example; (3) fictional example; (4) debate, (5) role play; (6) lecture; (7) class discussions/reflections; (8) small group discussions/reflections; (9) peer-led discussions; (10) individual reflections; (11) discussion posts; (12) simulations; (13) group collaborations; (14) games; (15) situated learning; (16) media example(s); (17) crash courses; (18) compare/contrast ethical guidelines; (19) hands-on activity/applied AI; and (20) other.

Learning Outcomes Evaluation Method. Learning outcomes evaluation response options

include: (1) survey; (2) observation; (3) assignment; (4) interview with students; (5) student reflections; (6) focus groups; (7) other; (8) not reported; and (9) not applicable.

Learning Outcomes Analysis Method. Possible approaches to analyzing learning outcomes include: (1) thematic analysis; (2) content analysis; (3) discourse analysis; (4) descriptive statistics; (5) other; and (6) none.

Reported Learning Outcomes. Potential reported learning outcomes include: (1) critical thinking skills (oral); (2) critical thinking skills (written); (3) ethical thinking; (4) empathy; (5) knowledge of ethical principles; (6) ability to evaluate the social impact of systems; (7) raising awareness of ethical concerns; (8) other; and (9) not reported.

Outcome Direction of Change. If researchers report on the direction of change for study outcomes, this will be charted as follows: (1) increase; (2) decrease; (3) mixed changes; (4) no change; (5) other; (6) not reported; and (7) not applicable.

Step 5: Collating, summarizing, and reporting the results

Building on the data charting phase, the team will summarize the data according to each of the data charting elements. We will report on which elements have received the most and the least attention in this field of study, including specific examples of articles to illustrate what each data charting element represents. Following this, we will summarize recommended next steps to improve instructional design for teaching AI ethics in postsecondary contexts.

Discussion

Results will include an up-to-date synthesis of the current state of AI ethics education in postsecondary curricula, evaluated teaching strategies, and potential outcomes associated with AI

ethics education. This review will inform future research, policy development, and teaching practices, offering valuable insights for educators, policymakers, and researchers working towards responsible AI integration. Findings will contribute to enhanced understandings of the complexities of AI ethics education and have the potential to shape the ways trainees in multiple disciplines learn about the ethical dimensions of AI in practice.

Acknowledgements

The [funder *anonymized*], [funding reference number *anonymized*] provided student scholarship funding. The funders had no role in study design, in collection or interpretation of data, in writing the report, or in the decision to submit the article for publication.

References

1. Moor J. The Dartmouth College artificial intelligence conference: the next fifty years. AI Mag. 2006;27(4):87-91. doi: 10.1609/aimag.v27i4.1911
2. Pauketat JVT, Ladak A, Anthis JR. Artificial intelligence, morality, and sentience (AIMS) survey: 2023 update. PsyArXiv [preprint]. 2023 Sept 8 [cited 2025 April 20]: [7 p.]. Available from https://doi.org/10.31234/osf.io/9xsav
3. Nguyen D, Hekman E. The news framing of artificial intelligence: a critical exploration of how media discourses make sense of automation. AI Soc. 2024;39:437-51. doi: 10.1007/s00146-022-01511-1
4. Roe J, Perkins M. 'What they're not telling you about ChatGPT': exploring the discourse of AI in UK news media headlines. Humanit Soc Sci Commun. 2023;10(1):753. doi:

10.1057/s41599-023-02282-w

5. European Commission. Hiroshima process international code of conduct for organizations developing advanced AI systems. 2023 Oct 30 [cited 2025 April 20]: [8 p.]. Available from https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems
6. Buruk B, Ekmekci PE, Arda B. A critical perspective on guidelines for responsible and trustworthy artificial intelligence. Med Health Care Philos. 2020;23(3):387-99. doi: 10.1007/s11019-020-09948-1
7. Turner Lee N. Detecting racial bias in algorithms and machine learning. J Inf Commun Ethics Soc. 2018;16(3):252-60. doi: 10.1108/JICES-06-2018-0056
8. Shen L, Chen I, Grey A, Su A. Teaching and learning with artificial intelligence. In: Verma S, Tomar P, editors. Impact of AI technologies on teaching, learning, and research in higher education. Hershey: IGI Global; 2021. pp. 73-98. doi: 10.4018/978-1-7998-4763-2.ch005
9. Sowmia KR, Poonkuzhali S. Artificial intelligence in the field of education: a systematic study of artificial intelligence impact on safe teaching learning process with digital technology. J Green Eng. 2020;10(4):1566-83.
10. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res. 2021;23(4):e25759. doi: 10.2196/25759
11. Borenstein J, Howard A. Emerging challenges in AI and the need for AI ethics education. AI Ethics. 2021;1(1):61-5. doi: 10.1007/s43681-020-00002-7
12. Raji ID, Scheuerman MK, Amironesei R. You can't sit with us: exclusionary pedagogy in AI ethics education. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21); 2021 Mar 3-10. Virtual Event, Canada.

doi: 10.1145/3442188.3445914

13. Fiesler C, Garrett N, Beard N. What do we teach when we teach tech ethics?: a syllabi analysis. Proceedings of the 51st ACM Technical Symposium on Computer Science Education; 2020 Mar 11-14. Portland, USA. doi: 10.1145/3328778.3366825
14. Gorur R, Hoon L, Kowal E. Computer science ethics education in Australia – a work in progress. Proceedings of the 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE); 2020 Dec 8-11. Takamatsu, Japan. https://ieeexplore.ieee.org/document/9368375/
15. Suárez JL, Varona D. The ethical skills we are not teaching: an evaluation of university level courses on artificial intelligence, ethics, and society. CulturePlex Lab, Western University. 2021 [cited 2025 April 20]: [36 p.]. Available from https://cultureplex.ca/wp-content/uploads/2022/01/Ethical_Skills_We_Are_Not_Teaching_Report.pdf
16. Javed RT, Nasir O, Borit M, Vanhée L, Zea E, Gupta S, et al. Get out of the BAG! Silos in AI ethics education: unsupervised topic modeling analysis of global AI curricula. J Artif Intell Res. 2022;73:933-65. doi: 10.1613/jair.1.13550
17. Shilton K, Heidenblad D, Porter A, Winter S, Kendig M. Role-playing computer ethics: designing and evaluating the privacy by design (PbD) simulation. Sci Eng Ethics. 2020;26(6):2911-26. doi: 10.1007/s11948-020-00250-0
18. Shapiro BR, Lovegall E, Meng A, Borenstein J, Zegura E. Using role-play to scale the integration of ethics across the computer science curriculum. Proceedings of the 52nd ACM Technical Symposium on Computer Science Education; 2021 Mar 13-20. Virtual Event, USA. doi: 10.1145/3408877.3432525
19. Ferreira R, Vardi MY. Deep tech ethics: an approach to teaching social justice in computer

science. Proceedings of the 52nd ACM Technical Symposium on Computer Science Education; 2021 Mar 13-20. Virtual Event, USA. doi: 10.1145/3408877.3432449

20. Garrett N, Beard N, Fiesler C. More than "if time allows": the role of ethics in AI education. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society; 2020 Feb 7-9. New York, USA. doi: 10.1145/3375627.3375868
21. Weidener L, Fischer M. Artificial intelligence teaching as part of medical education: qualitative analysis of expert interviews. JMIR Med Educ. 2023;9:e46428. doi: 10.2196/46428
22. Tuovinen L, Rohunen A. Teaching AI ethics to engineering students: reflections on syllabus design and teaching methods. Proceedings of the Conference on Technology Ethics; 2021 Oct 20-22. Turku, Finland. https://ceur-ws.org/Vol-3069/FP_02.pdf
23. Fiesler C, Friske M, Garrett N, Muzny F, Smith JJ, Zietz J. Integrating ethics into introductory programming classes. Proceedings of the 52nd ACM Technical Symposium on Computer Science Education; 2021 Mar 13-20. Virtual Event, USA. doi: 10.1145/3408877.3432510
24. McDonald N, Akinsiku A, Hunter-Cevera J, Sanchez M, Kephart K, Berczynski M, et al. Responsible computing: a longitudinal study of a peer-led ethics learning framework. ACM Trans Comput Educ. 2022;22(4):1-21. doi: 10.1145/3469130
25. Shih PK, Lin CH, Wu LY, Yu CC. Learning ethics in AI—teaching non-engineering undergraduates through situated learning. Sustainability. 2021;13(7):3718. doi: 10.3390/su13073718
26. Munn L. The uselessness of AI ethics. AI Ethics. 2023;3:869-77. doi: 10.1007/s43681-022-00209-w

27. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. 2005;8(1):19-32. doi: 10.1080/1364557032000119616

28. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1. doi: 10.1186/2046-4053-4-1

29. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467-73. doi: 10.7326/M18-0850

30. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1(9):389-99. doi: 10.1038/s42256-019-0088-2

31. Cheng L, Varshney KR, Liu H. Socially responsible AI algorithms: issues, purposes, and challenges. J Artif Intell Res. 2021;71:1137-81. doi: 10.48550/arXiv.2101.02032

32. Dignum V. Responsible artificial intelligence: designing AI for human values. ITU J (Geneva). 2017;1:1-8. https://www.itu.int/en/journal/001/Documents/itu2017-1.pdf

33. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems. 1st ed. IEEE, 2019. [cited 2025 April 20]: [291 p.]. Available from https://sagroups.ieee.org/global-initiative/wp-content/uploads/sites/542/2023/01/ead1e.pdf

34. Richardson B, Gilbert JE. A framework for fairness: a systematic review of existing fair AI solutions. arXiv [preprint]. 2021 [cited 2025 April 20]: [28 p.]. Available from https://arxiv.org/abs/2112.05700

35. Hajian S, Bonchi F, Castillo C. Algorithmic bias: from discrimination discovery to fairness-aware data mining. Proceedings of the 22nd ACM SIGKDD International Conference on

Knowledge Discovery and Data Mining; 2016 Aug 13-17. San Francisco, USA. doi: 10.1145/2939672.2945386

36. Venkatasubramanian S. Algorithmic fairness: measures, methods and representations. Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems; 2019 Jun 30-Jul 5. Amsterdam, Netherlands. doi: 10.1145/3294052.3322192

37. Hanif A, Zhang X, Wood S. A survey on explainable artificial intelligence techniques and challenges. Proceedings of the IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW); 2021 Oct 25-29. Gold Coast, Australia. doi: 10.1109/EDOCW52865.2021.00036

38. Hamel C, Hersi M, Kelly SE, Tricco AC, Straus S, Wells G, et al. Guidance for using artificial intelligence for title and abstract screening while conducting knowledge syntheses. BMC Med Res Methodol. 2021;21:285. doi: 10.1186/s12874-021-01451-2

39. Cormack G, Grossman M. Scalability of continuous active learning for reliable high-recall text classification. Proceedings of the 25th ACM International Conference on Information and Knowledge Management; 2016 Oct 24-28. Indianapolis, USA. doi: 10.1145/2983323.2983776

40. Thomas J, Grazioso S, Brunton J, Ghouze Z, O'Driscoll P, Bond M, et al. EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis. University College London: EPPI Centre, UCL Social Research Institute. 2023 [cited 2025 Apr 20]. Available from https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2967

41. Katznelson G, Gerke S. The need for health AI ethics in medical school education. Adv Health Sci Educ Theory Pract. 2021;26:1447-58. doi: 10.1007/s10459-021-10040-3