Information Processing and Management 63 (2026) 104767

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

AI ethics education: A scoping review of pedagogy, curriculum, and assessment

Calvin Hillis [a,*], Maushumi Bhattacharjee [b], Batool AlMousawi [c], Riley Martens [c], Tarik Eltanahy [a], Sara Ono [a], Marcus Hui [d], Ba' Pham [e], Michelle Swab [f], Gordon V. Cormack [g], Maura R. Grossman [g], Ebrahim Bagheri [h], Zack Marshall [c]

[a] The Creative School, Toronto Metropolitan University, 350 Victoria St, Toronto, ON M5B 2K3, Canada
[b] Faculty of Law, McGill University, 845 Sherbrooke St W, Montreal, QC H3A 0G4, Canada
[c] Department of Community Health Sciences, University of Calgary, 2500 University Dr NW, Calgary, AB T2N 1N4, Canada
[d] Department of Biology, Queen's University, 99 University Ave, Kingston, ON K7L 3N6, Canada
[e] Institute of Health Policy, Management and Evaluation, University of Toronto, 27 King's College Cir, Toronto, ON M5S 1A1, Canada
[f] Faculty of Medicine, Memorial University, St. John's, NL A1C 5S7, Canada
[g] School of Computer Science, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada
[h] Faculty of Information, University of Toronto, 140St George St, Toronto, ON M5S 3G6, Canada

ARTICLE INFO

Keywords:
Artificial intelligence
AI ethics
Ethics education
Higher education
Pedagogy
Assessment
Scoping review

ABSTRACT

Background: Artificial intelligence (AI) is increasingly embedded in social and institutional decision-making, creating demand for ethically literate practitioners. Universities have responded by introducing AI ethics instruction, but the structure, content, pedagogy, and evaluation of these efforts remain unevenly documented.
Objective: To map and synthesize research on university level AI ethics education by characterizing course design, pedagogy, ethical themes, and assessment methods, and identifying evidence gaps that limit knowledge consolidation and instructional refinement.
Methods: We conducted a scoping review using Continuous Active Learning to screen 50,766 records up to 2024. 43 studies met inclusion after title, abstract, and full text review. We coded instructional design, curricular themes, pedagogical methods, and evaluation approaches using descriptive frequency counts and qualitative synthesis.
Results: Most included studies were conceptual or descriptive, with relatively few empirical evaluations. Instruction was concentrated in computing and engineering and primarily targeted undergraduate learners. Ethics content was more often embedded within technical courses than delivered as standalone offerings. Reported pedagogy relied heavily on lecture and case-based discussion, with fewer studies describing participatory formats such as simulations or role-play. Curricular emphasis clustered around bias/fairness and privacy, with comparatively less attention to governance, explainability, and trust. Evaluation most often relied on self-report and reflective methods, while validated instruments and performance-based assessments were less common, and behavioral or applied outcomes were rarely assessed.

* Corresponding author.
E-mail addresses: cthillis@torontomu.ca (C. Hillis), maushumi.bhattacharjee@mail.mcgill.ca (M. Bhattacharjee), batool.almousawi@ucalgary.ca (B. AlMousawi), riley.martens@ucalgary.ca (R. Martens), tarik.eltanahy@torontomu.ca (T. Eltanahy), sara.ono@torontomu.ca (S. Ono), marcushui87@gmail.com (M. Hui), ba.pham@theta.utoronto.ca (B. Pham), mswab@mun.ca (M. Swab), gvcormack@uwaterloo.ca (G.V. Cormack), maura.grossman@uwaterloo.ca (M.R. Grossman), ebrahim.bagheri@utoronto.ca (E. Bagheri), zack.marshall@ucalgary.ca (Z. Marshall).

https://doi.org/10.1016/j.ipm.2026.104767
Received 1 November 2025; Received in revised form 19 February 2026; Accepted 21 March 2026
Available online 27 March 2026

Conclusions: The literature suggests a field oriented toward awareness-building more than measurable ethical competence. Clearer competency claims, stronger assessment transparency, and greater alignment between instructional design and evaluation would improve comparability across studies and support evidence-informed course development.

1. Introduction

Artificial intelligence (AI) is now a central feature of modern life. Across domains such as healthcare, education, governance, and communication, AI systems influence decisions that affect individuals and communities (Shen et al., 2021; Sowmia & Poonkuzhali, 2020; Yin et al., 2021). These systems offer opportunities for efficiency, innovation, and discovery, yet they also introduce ethical challenges related to bias, transparency, accountability, and privacy (Buruk et al., 2020; Turner Lee, 2018). As AI technologies continue to evolve and become more autonomous, the responsibility for designing and managing them in ways that promote fairness and public trust has become a critical societal concern.

Universities play an essential role in responding to this challenge. These institutions educate the engineers, scientists, and professionals who will develop, deploy, and regulate AI systems. Preparing these students requires more than technical expertise. It also demands ethical literacy, critical thinking, and an awareness of the social consequences of technological decisions (Borenstein & Howard, 2021; Raji et al., 2021). As a result, educators are being called to integrate ethics into AI curricula so that future practitioners are better equipped to anticipate harm, ensure accountability, and align design choices with human values (Borenstein & Howard, 2021).

1.1. Rationale for the review

Although there is growing consensus on the need for AI ethics education, published research describing how AI ethics education is implemented and evaluated in higher education remains heterogeneous and dispersed across disciplines and venues, with substantial variation in course design and reporting (Wiese et al., 2025; Brown et al., 2024). Individual universities have introduced ethics modules within computing programs or developed stand-alone courses focused on responsible AI. In higher education computing curricula, embedded-ethics models illustrate how ethical reasoning can be integrated within existing technical courses rather than treated exclusively as separate content (Grosz et al., 2019). Across published accounts, initiatives vary in design, scope, and evaluation focus. Some emphasize fairness and bias, others focus on privacy or governance, and evaluation frequently relies on student self-report, course evaluations, or instructor reflections rather than direct measures of learning outcomes (Brown et al., 2024).

Several recent reviews have begun to synthesize this emerging evidence base. Wiese et al. (2025) conducted a systematic literature review examining how ethics is taught within AI education contexts, focusing primarily on computing courses and identifying gaps in assessment practices. Brown et al. (2024) reviewed ethics instruction across computing education more broadly, documenting instructional strategies and highlighting inconsistencies in outcome evaluation. While these reviews have mapped portions of the landscape, important gaps remain. Existing reviews have focused predominantly on computing contexts, leaving interdisciplinary approaches less well characterized. Moreover, the mapping of how specific ethical principles such as fairness, accountability, transparency, explainability, and governance are addressed in curricula has not been synthesized comprehensively across the full range of publication types, including conceptual frameworks, course descriptions, and empirical evaluations.

This scoping review addresses these gaps by providing a comprehensive map of AI ethics education research that includes both computing and non-computing contexts, spans multiple publication types (empirical studies, conceptual papers, frameworks, and course descriptions), and explicitly codes for coverage of key ethical principles. Unlike prior systematic reviews that focused on identifying best practices or effect sizes, this scoping review maps the breadth and characteristics of the evidence base to identify what has been studied, how it has been reported, and where substantial knowledge gaps persist. This descriptive approach is particularly appropriate given the heterogeneity of the literature and the nascent state of evaluation practices in this domain Peters et al. (2020). By consolidating this diverse literature, the review provides educators with a structured overview of reported course models, instructional strategies, and evaluation approaches, and identifies priority areas for future empirical research and methodological development.

1.2. Purpose and objectives

The purpose of this scoping review is to map and synthesize research on the teaching and assessment of AI ethics in higher education and to identify gaps in curricula, pedagogy, and assessment practices.

The specific objectives of the review are to

1. Describe how AI ethics education is reported in higher education, including disciplinary contexts and publication formats.
2. Characterize how AI ethics education is delivered, including course structure and pedagogical approaches
3. Synthesize the AI ethics topics and curricular themes addressed across studies.
4. Identify how student learning is assessed and where evidence gaps persist, including the types and reported domains of learning outcomes.

To guide this inquiry, two research questions were formulated. The first asks how universities design and deliver AI ethics education across disciplines and delivery modes. The second asks how courses and programs evaluate student learning and behavioural outcomes in ways that reflect the social and ethical expectations of AI practice. These objectives and research questions directly motivate the contributions described in Section 1.4.

Table 1 summarizes the alignment among the review objectives, research questions, and findings sections.

1.3. Scope and conceptual framework

This review focuses on university education at the undergraduate, and graduate levels, encompassing both technical and non-technical disciplines. Artificial intelligence is defined broadly to include machine learning, deep learning, natural language processing, and related computational methods that influence or automate decision making. AI ethics refers to the principles, frameworks, and practices that guide the responsible design, deployment, and governance of such systems.

The conceptual foundation of this review builds on internationally recognized frameworks and standards for ethical artificial intelligence, including guidance from the IEEE, the European Commission, the OECD, UNESCO, and the National Institute of Standards and Technology, and on comparative syntheses that map areas of consensus across prominent AI ethics guidelines (Jobin et al., 2019; Fjeld et al., 2020; European Commission, 2019; OECD, 2019; UNESCO, 2021; NIST, 2023). These syntheses report convergence across prominent AI ethics guidelines around recurring themes including fairness, accountability, transparency and explainability, responsibility, privacy, and safety. These themes shape how this review interprets the pedagogical and evaluative strategies described in the literature (Jobin et al., 2019; Fjeld et al., 2020). In addition, this foundation is informed by work that synthesizes ethical principles and recommendations for a "Good AI Society" (Floridi et al., 2018) and by critiques emphasizing that shared high-level principles are insufficient without institutional mechanisms, professional norms, and governance arrangements that translate principles into practice (Mittelstadt, 2019). Complementing these frameworks and critiques, Morley et al. (2020) catalog AI ethics tools and methods for translating principles into practice, aligning with this review's focus on how ethical themes are operationalized and evaluated in educational settings. To clarify how this review defines these ethical themes and applies them to interpret the included studies, we draw on prior literature as follows.

Accountability concerns the distribution of responsibility for the impacts of AI systems across their entire lifecycle, from data collection and model training to deployment and use (Dignum, 2017). Initiatives such as the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2019) regard accountability as a central expectation for ethical decision making and governance (2019). Algorithmic bias refers to unintended and often discriminatory outcomes that arise from historical patterns within training data or from proxy variables correlated with sensitive attributes (Richardson & Gilbert, 2021). Mitigating bias requires deliberate, discrimination-aware design and continuous evaluation attentive to social contexts and population differences. Closely related is algorithmic fairness, which involves the design of methods and metrics that promote equitable outcomes across groups and settings. Fairness is inherently socio-technical and draws on insights from computer science, law, philosophy, economics, and the social sciences to balance quantitative criteria with broader conceptions of justice and equity (Richardson & Gilbert, 2021; Hajian et al., 2016; Venkatasubramanian, 2019). Transparency and explainability focus on the openness and interpretability of AI systems. Transparency relates to how systems communicate their purpose and limitations to users and interest holders, while explainability concerns the ability to understand how models produce outcomes so that claims about reliability and fairness can be examined and challenged (Cheng et al., 2021; Hanif et al., 2021). Together these principles provide a coherent basis for understanding the ethical expectations that underlie AI education and for assessing how curricula cultivate students' capacity to engage with them critically and responsibly.

The review includes English language studies from all regions and disciplines to capture the diversity of AI ethics education worldwide. It also considers different instructional modalities such as lectures, seminars, project-based learning, and online formats. The inclusion of interdisciplinary programs reflects the understanding that ethical reasoning about AI extends beyond computer science to fields such as health, law, business, and the humanities and social sciences.

Table 1
Review objectives and research questions.

Review objective	Research question(s)	Addressed in Findings
Identify how AI ethics is taught across different disciplinary contexts and instructional formats	RQ1: How do institutions design and deliver AI ethics education across disciplines and delivery modes.	3.3 Course Structure and Delivery; 3.5 Learning Designs; 3.8 Population
Examine the methods used to assess student learning outcomes in AI ethics education.	RQ2: How courses and programs evaluate student learning and behavioural outcomes in ways that reflect the social and ethical expectations of AI in practice.	3.6 Evaluation Methods and Analytical Practices; 3.7 Reported Learning Outcomes
Evaluate the extent to which current educational approaches address fairness, accountability, transparency, explainability, and governance.	RQ1: How do institutions design and deliver AI ethics education across disciplines and delivery modes.	3.4 Curriculum Themes
Highlight areas where evidence is insufficient and propose directions for future pedagogical and empirical research.	Supports both RQ1 and RQ2	3.2 Characteristics of Included Studies, 3.6 Evaluation Methods and Analytical Practices, 3.9 Synthesis

1.4. Significance of the study

This review advances understanding of how AI ethics is conceptualized, taught, and evaluated in university education by synthesizing a multidisciplinary evidence base and providing three concrete contributions that extend prior reviews. First, it provides a comprehensive descriptive map of AI ethics education research by consolidating scattered literature into an organized synthesis of course contexts, instructional formats, and ethics-related content emphases. While existing reviews (Wiese et al., 2025; Brown et al., 2024) have synthesized empirical studies of ethics in computing education, this scoping review broadens the evidence base by including conceptual frameworks, pedagogical proposals, and course descriptions alongside empirical evaluations, and by including studies across disciplines. This inclusive approach provides a more complete picture of how the field is being developed across publication types and disciplinary contexts, even though computing and engineering contexts remain predominant in the literature.

Second, it provides a systematic synthesis of reported evaluation practices by identifying what outcomes are assessed, what instruments or evidence sources are used, and where reporting is insufficient to support claims about learning effects. By coding evaluation methods across diverse publication types, including conceptual papers and pedagogical proposals alongside empirical evaluations, the review reveals patterns in how learning is conceptualized and measured, and identifies specific gaps in reporting rigor and methodological transparency. The synthesis documents widespread reliance on self-report measures and identifies instances where learning outcomes are claimed without supporting empirical evidence.

Third, it translates these mapped findings into a targeted research agenda that prioritizes clearer outcome specification, improved reporting of analysis procedures, explicit connections between pedagogical strategies and ethical principles, and greater cross-study comparability. This agenda is grounded in the patterns documented in the review and is designed to support both researchers planning future studies and educators seeking evidence to inform curricular decisions.

By consolidating research evidence on AI ethics education, this review provides educators, program designers, and policymakers with a structured knowledge base for designing curricula that prepare students to engage critically and responsibly with the ethical dimensions of AI systems. It responds to institutional needs for evidence-informed program development while identifying where future research can strengthen the empirical foundation for AI ethics pedagogy.

The remainder of the paper reports the review method and screening process, presents descriptive findings on institutional design and evaluation practices, and then interprets implications for research and university curriculum development in relation to the mapped evidence.2 Methods

1.5. Overview and design

This review followed a systematic and transparent approach to identify, select, and synthesize studies on the teaching and evaluation of AI ethics in higher education. The process was informed by the methodological framework of Arksey and O'Malley for scoping reviews and was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) (Arksey & O'Malley, 2005; Tricco et al., 2018). The review protocol was developed prospectively and defined all stages of the process, including the research questions, search strategy, inclusion and exclusion criteria, screening procedures, and data charting framework (Hillis et al., 2025). Each step was documented to ensure methodological transparency and reproducibility. A summary of the overall process is shown in Fig. 1, which presents the PRISMA flow diagram indicating the number of records identified, screened, assessed for eligibility, and included.

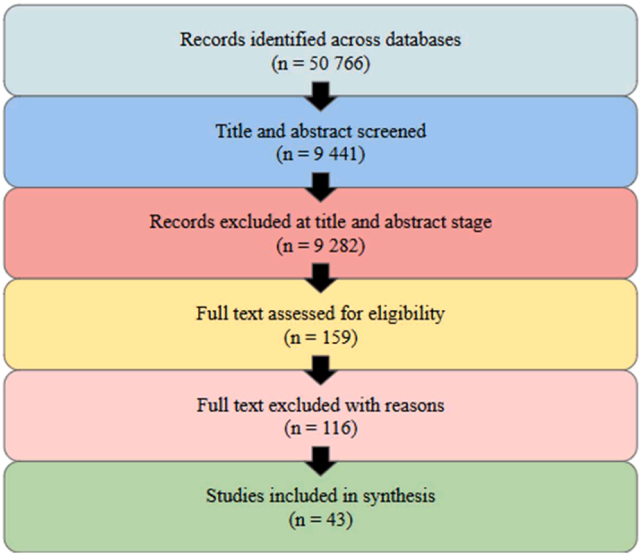

Fig. 1. PRISMA flow diagram of the study selection process.

A scoping review methodology was selected because the aims of this study are to map and characterize the extent, characteristics, and evaluation practices of AI ethics education research across heterogeneous publication types and study designs, rather than to estimate pooled effects or determine intervention efficacy. Scoping reviews are recommended when a body of literature is conceptually diffuse, methodologically diverse, and includes a substantial proportion of descriptive and non-intervention evidence, and when the objective is to identify how a field has been studied and where evidence gaps limit cumulative knowledge (Arksey & O'Malley, 2005; Levac et al., 2010; Peters et al., 2020). AI ethics education is a developing area of scholarship characterized by conceptual diversity and methodological variability. In this domain, the evidence base includes conceptual and descriptive accounts alongside a smaller set of empirical evaluations, and outcomes are often reported using non-comparable measures, making meta-analysis inappropriate and a systematic review focused on effect estimates premature. The scoping approach therefore aligns with this review's objectives by enabling comprehensive coverage and structured categorization of learning designs, curricular themes, and evaluation strategies, while also identifying limitations in the empirical evidence that warrant future research and more evaluative synthesis.

1.6. Information sources and search strategy

The search strategy was designed in collaboration with an academic librarian with expertise in scoping review methodology. Searches combined controlled vocabulary and free-text keywords related to education, ethics, and artificial intelligence. Eight bibliographic databases were searched with no date limiters, and the full search strings for each database are provided in Appendix B. 2011 corresponds to the publication year of the earliest included study. A summary of databases, coverage periods, and key concepts is presented in Table 2.

To ensure comprehensive coverage across the disciplinary domains in which AI ethics education is published, the search strategy combined databases selected for their complementary scope. PubMed (MEDLINE) and Embase were included to capture biomedical, clinical, and health-professions education literature. Scopus was included as a multidisciplinary index that supports broad retrieval across the social sciences, engineering, and health research. ERIC was included to capture education-focused scholarship, including pedagogical and curriculum studies that may not be indexed in biomedical databases. APA PsycInfo was included to capture psychology and learning sciences work relevant to ethics learning, professional decision-making, and educational evaluation. IEEE Xplore and the ACM Digital Library were included to capture computer science and engineering venues. Library, Information Science & Technology Abstracts (LISTA) was included to capture information science scholarship relevant to information ethics, information policy, and educational interventions reported in library and information science venues. Finally, we searched ProQuest Dissertations and Theses to identify relevant doctoral and master's work that may contain early pedagogical innovations, curricular designs, or evaluation studies that are not yet represented in the peer-reviewed journal literature, which is consistent with guidance that comprehensive searches may include grey literature sources such as dissertations and theses (Lefebvre et al., 2025).

This database-selection approach aligns with common scoping-review practice in other interdisciplinary domains, where authors combine discipline-specific databases with broad indexing services to capture breadth. For example, Brown et al. conducted a scoping review in an applied accessibility and design domain and searched a mixed set of databases spanning health, psychology, education, and multidisciplinary indexing (MEDLINE, PsycINFO, CINAHL, Embase, ERIC, and Scopus), illustrating a deliberate strategy to capture a topic distributed across multiple research traditions (Brown et al., 2021). Similarly, Sarraf-Yazdi et al. examined professional identity formation in undergraduate medical education and searched PubMed, Embase, PsycINFO, ERIC, and Scopus, again combining biomedical, psychology, education, and multidisciplinary sources to capture the topic's full scope (Sarraf-Yazdi et al., 2021). Other scoping reviews demonstrate comparable rationales for incorporating computing and engineering repositories alongside broad indexes when the topic intersects with technical research dissemination norms, including the inclusion of ACM Digital Library and IEEE Xplore within larger multi-database search strategies (Chan et al., 2024; Mevlevioğlu et al., 2023; Braz & Martín del Pozo, 2025). Finally, scoping review methods are also frequently extended beyond conventional academic indexing when the evidence base is distributed across publication types, including nonacademic sources, reinforcing the appropriateness of supplementing database searches with strategies intended to capture work outside standard journal channels (Francombe et al., 2022).

1.7. Eligibility criteria

Eligibility was established using the Population–Concept–Context framework. Studies were included or excluded based on the

Table 2
Summary of databases searched and coverage.

Database	Coverage period	Date searched	Records retrieved
PubMed (MEDLINE)	Since Database Inception–2024	26 June 2024	9753
Embase	–2024	26 June 2024	11,724
Scopus	–2024	26 June 2024	25,151
ERIC	Since Database Inception–2024	26 June 2024	421
LISTA	Since Database Inception–2024	27 June 2024	561
APA PsycInfo	Since Database Inception–2024	27 June 2024	1737
IEEE Xplore	Since Database Inception–2024	28 August 2024	682
ACM Digital Library	Since Database Inception–2024	28 August 2024	235
ProQuest Dissertations and Theses	Since Database Inception–2024	26 June 2024	502

criteria summarized in Table 3.

1.8. Screening and selection of studies

All retrieved records were exported to a reference manager and uploaded into a Continuous Active Learning (CAL®) platform to prioritize title and abstract screening (Cormack & Grossman, 2016). CAL uses active machine learning to iteratively re-rank unscreened records based on reviewers' include and exclude decisions so that records most likely to meet eligibility criteria are screened earlier (Norman et al., 2019). Because prioritization changes the order of screening, review teams must specify a transparent stopping approach for when additional screening is unlikely to yield eligible studies (Callaghan & Müller-Hansen, 2020; Hamel et al., 2021). We applied a developer recommended pre-specified low-yield stopping criterion and ended title and abstract screening once five consecutive batches of 100 records yielded fewer than 5% newly included records, indicating diminishing returns and likely saturation. Full texts of potentially eligible studies were then reviewed in duplicate.

Two reviewers independently screened titles and abstracts using the eligibility criteria in Table 2. Disagreements were resolved through discussion, with a third reviewer available to mediate unresolved cases. In total, 50,766 records were identified across databases and 9441 records were screened at title and abstract. Full texts of potentially eligible studies were then reviewed in duplicate. Of the 159 full texts assessed, 116 were excluded because AI ethics education was only briefly mentioned ($n = 52$) or because AI ethics education was a small or insignificant component of the paper ($n = 64$). Ultimately, 43 studies were included, restricted to papers in which AI ethics education was a primary focus. The PRISMA flow diagram in Fig. 1 summarizes the selection process and reports the full-text exclusion reasons.

1.9. Data charting and extraction

Data extraction followed a structured process to ensure accuracy and comparability. A standardized charting form was developed and piloted on a subset of studies before full extraction. One reviewer extracted data and a second reviewer verified all entries. The charting framework is summarized in Table 4.

Descriptive statistics were used to summarize study characteristics and coded variables using frequency counts (n) and percentages (%), including distributions across publication years, disciplines, and educational levels. No inferential statistical tests were conducted, consistent with scoping review guidance that emphasizes mapping and characterizing heterogeneous evidence bases rather than estimating pooled effects (Arksey & O'Malley, 2005; Tricco et al., 2018; Peters et al., 2020). Qualitative data on pedagogy, assessment, and outcomes were analyzed thematically. Given heterogeneity in study designs and measures, quantitative meta-analysis was not feasible. Findings were organized around pedagogical practices and evaluation approaches.

All records of the search, screening decisions, and coding were maintained in a dedicated database (EPPI-Reviewer V6). The dataset, including extracted variables and verbatim search strategies, is available in the supplementary materials. Maintaining a complete audit trail enhances the transparency and reproducibility of the review.

2. Findings

The final corpus reveals a field still defining its theoretical and pedagogical boundaries. Scholarship on AI ethics education has expanded rapidly since 2021, reflecting institutional, professional, and societal interest in embedding ethics within AI-related programs. Yet the body of work remains methodologically diverse and conceptually unsettled. Most studies rely on descriptive accounts or narrative reflections rather than empirical evaluations. This pattern suggests that AI ethics education is still in its formative stage, where the community is building consensus on what ethical competence entails and how it can be fostered through curriculum and pedagogy. Table 5 summarizes the characteristics of all included studies. Results from individual sources of evidence are found in Appendix C.

Table 3
Inclusion and exclusion criteria based on the population–concept–context framework.

Element	Inclusion criteria	Exclusion criteria
Population	University learners and educators including both undergraduate and graduate levels	Primary or secondary learners or professional training that is not situated in formal higher education
Concept	Teaching and evaluation of AI ethics including curriculum design, course content, pedagogical strategy, or assessment of learning	Studies addressing general computer ethics without explicit AI content or lacking pedagogical detail or where AI ethics education is not a primary focus.
Context	Formal higher education delivered in any format including classroom, online, or hybrid	Non-educational or corporate settings
Language and date	English publications from Inception to July 2024	Non-English publications
Publication type	Peer-reviewed journal articles, conference papers, dissertations, and book chapters reporting educational interventions or evaluations	Editorials, commentaries, or opinion pieces

Table 4
Data extraction and coding framework.

Category	Data extracted	Description or coding example
Study identification	Author, year, publication type	Bibliographic information
Educational level	Undergraduate, graduate	Level of learner or program
Disciplinary context	Computer science (broadly), non-computer science (broadly)	Academic discipline
Format of report	Course plan, general course information, pedagogical review, presentation of pedagogical intervention	Level of content discussed
Course type	Embedded ethics, standalone course, content specific ethics	Structure and delivery mode
Ethical focus	Fairness, accountability, transparency, explainability, governance, bias, trust, responsibility, safety, privacy	Principal ethical themes
Learning Design	Lecture, project-based, case study, experiential learning	Instructional strategy
Evaluation method	Survey, reflection, interview, performance task, assignment grades, purported (learning outcome)	Assessment instrument
Outcome category	Knowledge, skill, attitude, behavior	Dimension of learning outcome
Direction of change	Positive, negative, mixed, not reported	Summary of outcome trend

Table 5
Characteristics of included studies - study-level data supporting each count are provided in Appendix C using Study IDs (S1–S43).

Category	Frequency	Description	Studies (IDs)
Conceptual essays	21	Theoretical or normative justification for AI ethics instruction	S2, S5–S10, S14, S17, S18, S23, S24, S26, S27, S32–S34, S37–S40
Detailed course plans	18	Structured curricula or syllabi with explicit objectives	S1, S4, S10, S12, S13, S15, S17, S19, S25, S27, S29, S33, S37–S39, S41–S43
Pedagogical frameworks	16	Model-based proposals for integrating ethics in AI education	S2, S5, S6, S16, S20, S24, S26, S27, S31, S33, S34, S35–S37, S40, S42
General course descriptions	12	Informal accounts lacking evaluation or assessment detail	S3, S7, S10, S11, S16, S17, S21, S22, S24–S26, S30
Review or mapping papers	7	Synthetic works summarizing prior efforts	S2, S14, S20, S28, S31, S32, S36

2.1. Study selection and distribution

From an initial pool of 50,766 records, 9441 underwent title and abstract screening. After full-text review of 159 articles, 43 studies met inclusion criteria. Publication activity increased sharply after 2021, coinciding with heightened policy and public discourse surrounding responsible AI. The excluded full-text papers frequently addressed ethics rhetorically without engaging pedagogical theory or assessment. This demonstrates that while ethical AI is widely discussed, structured and assessable educational practices remain limited. The included papers represent a diverse set of disciplines and course types, reflecting an emerging ecosystem of instructional experimentation rather than a consolidated pedagogical paradigm.

2.2. Characteristics of included studies

Table 5 shows that the corpus encompasses conceptual essays, course descriptions, theoretical proposals, and reviews. The predominance of conceptual and argumentative papers reflects an early emphasis on articulating rationales for AI ethics instruction before empirical validation. Only a minority of studies conducted systematic evaluations, underscoring the scarcity of formal assessment frameworks in this domain. This imbalance mirrors the developmental trajectory of earlier technology-ethics movements, where advocacy and conceptual justification precede methodological rigor. The limited number of review papers also signals the nascency of meta-pedagogical reflection within AI ethics education research.

Table 6
Pedagogical structures and approaches.

Theme	Frequency	Representative Focus	Studies (IDs)
Embedded ethics	23	Ethics instruction is integrated into technical modules	S1, S4, S7–S8, S10–S13, S18–S20, S22–S25, S29–S30, S32, S35–S38, S40
Stand-alone ethical content	11	Ethics instruction is located within its own pedagogical modules	S6–S7, S9, S17, S24, S26–S27, S33, S39, S41–S42
Content specific ethics	3	Ethics instruction is specified, for example data ethics	S3, S15, S22
Activity based	2	Ethics instruction centers around a specific activity	S17, S35

2.3. Course structure and delivery

Ethics instruction appears most frequently embedded within technical courses. Table 6 presents the distribution of pedagogical structures and approaches. For clarity, the term "pedagogical frameworks" in Table 5 refers to the format of the included publication, specifically studies that propose or describe a framework or model for integrating AI ethics into higher education curricula. In contrast, Table 6 summarizes the curricular placement and delivery structure of ethics instruction as reported in the included studies. These categories are not hierarchical. Table 5 characterizes the types of evidence included in the corpus, whereas Table 6 characterizes how ethics instruction is positioned and delivered within the reported educational settings.

Interpreting Table 6 through this lens highlights where ethics instruction is positioned, and also insights into why embedded approaches are frequently adopted in practice.

Embedding ethics in existing courses lowers administrative barriers and allows instructors to situate ethical reasoning within tangible design and coding practices. However, this approach often constrains depth, as ethics topics are compressed into brief modules or single assignments. Standalone courses, though less common, tend to foster more sustained philosophical reflection and critical dialogue. The dominance of embedded delivery suggests a pragmatic orientation toward curricular integration but reveals a need for stronger scaffolding that enables ethical inquiry beyond the level of awareness.

2.4. Curriculum content themes

As shown in Table 7, the most frequently taught topics are algorithmic bias and fairness, social impact and responsibility, and privacy. These themes align with widespread public concerns and are easily demonstrated through data-driven examples, making them pedagogically accessible. In contrast, topics such as explainability, governance, trust, and regulation appear far less often, despite their importance in responsible AI practice. The imbalance arises partly from the availability of illustrative case studies and teaching resources concentrated around bias and fairness. While these areas are vital, the lack of engagement with structural governance and policy literacy limits students' ability to navigate systemic ethical challenges. The field therefore remains oriented toward individual-level ethical sensitivity rather than institutional or regulatory awareness.

2.5. Learning designs

Table 8 details the instructional methods reported in the corpus. Lecture and case-based teaching dominate, reflecting continuity with traditional higher-education formats. Participatory methods such as design projects, role play, or simulations are less common, even though they align more closely with the experiential nature of ethical decision-making.

2.6. Evaluation methods and analytical practices

Evaluation practices are summarized in Table 9. Surveys and reflective assignments predominate, offering self-reported evidence of attitudinal change but rarely addressing skill development or behavioural application. Few studies used validated instruments for moral reasoning or ethical decision-making, and only a small subset described analytic rigor in qualitative evaluations. This reliance on descriptive or self-assessed outcomes limits confidence in claims of instructional effectiveness. It also reflects a wider challenge in educational research on ethics, where affective and cognitive dimensions are difficult to measure. Developing robust evaluative tools will be critical for advancing this area beyond anecdotal validation.

2.7. Reported learning outcomes

Reported outcomes are summarized in Table 10. The dominant trend involves cognitive and attitudinal gains where students report

Table 7
Curriculum and pedagogical themes.

Theme	Representative Focus	Frequency	Studies (IDs)
Foundational ethics and ethical reasoning	Broad ethical considerations, formal ethics, ethical concepts	31	S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S15, S16, S17, S20, S23, S24, S25, S26, S29, S30, S31, S33, S34, S35, S36, S37, S38, S39, S40, S41, S42
Bias and fairness	Equity and discrimination in data and algorithmic design	27	S1, S3, S5, S6, S8, S10, S12, S13, S15, S16, S17, S18, S19, S21, S22, S23, S26, S27, S29, S32, S33, S34, S35, S36, S37, S40, S43
Social impact, harms, and sustainability	Societal outcomes, harms, and future implications	18	S5, S9, S11, S17, S19, S24, S25, S27, S29, S30, S31, S33, S34, S36, S37, S39, S40, S42
Privacy and surveillance	Data governance and consent in AI use	15	S1, S4, S9, S11, S15, S18, S19, S26, S27, S32, S33, S34, S35, S39, S42
Transparency and explainability	Interpretability of models and decisions	10	S1, S5, S9, S23, S26, S27, S35, S36, S37, S42
Accountability	Role of developers and institutions	9	S1, S13, S15, S16, S23, S27, S33, S35, S36
Governance and regulation	Legal and institutional mechanisms	9	S1, S3, S15, S17, S24, S26, S30, S33, S39
Security and safety	Risk mitigation and harm prevention	9	S1, S11, S19, S26, S30, S34, S37, S39, S42
Trust and human oversight	Public confidence and human control	2	S3, S30

Table 8
Learning designs.

Method or Tool	Frequency	Observed Purpose	Studies (IDs)
Case study discussion	18	Application to real-world ethical scenarios	S1, S2, S3, S5, S7, S15, S16, S17, S18, S20, S23, S26, S30, S32, S33, S34, S36, S37
Lecture-based instruction	23	Foundational knowledge dissemination	S1, S3, S4, S9, S10, S12, S13, S15, S16, S21, S22, S23, S25, S26, S27, S29, S30, S33, S34, S38, S39, S41, S42
Reflection or journaling	11	Self-assessment and value articulation	S2, S4, S12, S17, S26, S31, S33, S35, S37, S41, S42
Hands-on project or lab	12	Contextualized ethical reasoning	S5, S13, S22, S26, S27, S29, S30, S37, S38, S41, S42, S43
Collaborative group work	8	Shared reasoning and debate	S1, S17, S18, S26, S27, S38, S41, S42
Role play or simulation	6	Experiential understanding of moral dilemmas	S1, S2, S17, S18, S19, S26

Table 9
Evaluation approaches.

Evaluation Approach	Frequency	Representative Focus	Studies (IDs)
Survey or questionnaire	15	Measuring attitudinal change	S15, S17, S18, S21, S22, S23, S25, S29, S33, S34, S36, S37, S38, S39, S43
Assignment or exam	6	Evaluating conceptual comprehension	S11, S12, S13, S15, S25, S36
Interviews with student	3	Qualitative insight into learning process	S12, S32, S38

Table 10
Reported learning outcomes.

Outcome Category	Frequency	Typical Evidence Reported	Studies (IDs)
Ethical awareness	28	Increased sensitivity to ethical issues	S1, S8, S10, S13, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S29, S30, S31, S32, S33, S34, S35, S36, S37, S38, S39, S41, S42
Ethical reasoning	20	Improved ability to analyze dilemmas	S7, S12, S16, S17, S19, S21, S22, S23, S24, S25, S29, S31, S33, S34, S35, S37, S38, S39, S40, S42
Social impact assessment	18	Better evaluation of AI's consequences	S8, S11, S12, S15, S17, S19, S22, S23, S24, S25, S29, S30, S31, S34, S36, S37, S41, S42
Understanding of principles	14	Enhanced knowledge of fairness, bias, and responsibility	S1, S6, S7, S10, S11, S12, S20, S23, S25, S30, S33, S34, S35, S39
Critical thinking	12	Strengthened analytic and reflective capacity	S1, S4, S19, S24, S30, S31, S33, S34, S35, S37, S41, S42

greater ethical awareness, improved conceptual understanding, and heightened recognition of social implications. These findings indicate successful sensitization but limited evidence of applied moral reasoning or behavioural change. Only a few studies linked learning to practical design contexts or used performance-based assessments. This trend highlights the difficulty of evaluating ethical competence that extends beyond classroom reflection. To move forward, programs will need to integrate longitudinal assessment and context-specific evaluation to capture durable learning effects.

2.8. Population

Population numbers of students referred to in the collected articles are summarized in Table 11. Across studies, participant pools

Table 11
Population.

Outcome Category	Frequency	Representative Focus	Study IDs
Computer Science / Engineering - Undergrad	26	Studies that reported participants who were undergraduate students in computer science and/or engineering (broadly)	S3, S4, S5, S6, S7, S8, S11, S12, S13, S16, S17, S18, S19, S20, S21, S23, S24, S25, S26, S29, S32, S33, S34, S36, S38, S39
Computer Science / Engineering - Graduate	17	Studies that reported participants who were graduate students in computer science and/or engineering (broadly)	S3, S4, S5, S8, S9, S10, S12, S15, S23, S25, S26, S27, S30, S32, S35, S41, S42
Non-computer science / Engineering - Undergrad	6	Studies that reported participants who were undergraduate students in non-computer-science disciplines (broadly)	S5, S24, S25, S33, S37, S43
Non-computer science / Engineering - Graduate	4	Studies that reported participants who were graduate students in non-computer-science disciplines (broadly)	S5, S25, S35, S43

skew toward technical cohorts and earlier stages of training. This narrows how AI ethics learning is conceptualized, centering developer perspectives and classroom reflection over the ethics of use, oversight, and sector-specific practice. The result is curricular siloing that limits transfer across contexts and obscures how outcomes vary by disciplinary culture, prior expertise, and career orientation. External validity is also constrained when populations mirror a single learner profile, making it difficult to infer what works in law, business, health, or public policy classrooms.

2.9. Synthesis

The reviewed studies depict AI ethics education as an expanding domain characterized by diverse instructional and evaluative approaches. Across the corpus, studies more often emphasize exposure to ethical issues and self-reported learning than performance-based assessment, with surveys and reflective assignments predominating and few studies using validated instruments for moral reasoning or ethical decision-making. Limited use of standardized assessment approaches makes it difficult to compare outcomes across studies. In addition, the literature remains concentrated in technical contexts and earlier stages of training, which constrains what can be inferred about how AI ethics education functions across disciplinary settings and professional programs.

Most studies originated from computer science and engineering departments, and undergraduate learners accounted for the majority of participants, with fewer initiatives reported at the graduate level. Fewer studies reported implementations in non-technical disciplines. This distribution indicates that the included literature disproportionately reflects AI ethics education as it is delivered within technical programs. Greater representation from additional disciplines could broaden the range of curricular approaches reported in the literature and support comparisons across different educational contexts.

3. Discussion

3.1. Course structure and content delivery

This review found a strong preference for embedding AI ethics education within technical courses, rather than teaching it as a standalone subject. Embedding ethics alongside technical content may help students encounter ethical issues in closer proximity to design and implementation contexts, but a key remaining concern is whether integration allows sufficient depth in ethical reasoning. Some articles suggested that embedded delivery can pique students' interest in ethics and improve self-perceived ethical awareness and decision-making (Grosz et al., 2019; Kopec et al., 2023). A related challenge is ensuring instructors have the cross-domain knowledge and shared vocabulary needed for interdisciplinary teaching (Grosz et al., 2019). Targeted inquiry is needed to assess the strengths and limitations of embedded ethics more rigorously.

Standalone courses may provide broader foundations in ethical theory but can risk abstraction if discussion is not connected to technical practice. This trade-off between breadth and contextual grounding may be mitigated by incorporating domain-specific principles, such as healthcare, which can anchor ethical reasoning in sector expectations that shape AI design and deployment (Guizzardi et al., 2023; Zuber et al., 2022).

Direct comparative studies of embedded versus standalone approaches in AI ethics education were not found. One relevant comparison was found outside of this review, in a study of postgraduate students enrolled in business ethics courses (Coldwell et al., 2020), which demonstrated that both approaches positively influenced students' moral reasoning and ethical decision-making, albeit in different ways. That study found no statistically significant difference in perceived effectiveness; qualitative responses suggested that standalone courses better developed theoretical reasoning, while embedded courses better supported practical ethical decision-making.

Future work should test these patterns within AI ethics education through comparative studies that use shared outcomes and analytic rubrics, track knowledge, skills, attitudes, and behaviors across time, while including interdisciplinary cohorts.

3.2. Curriculum content and themes

AI ethics education often centers on broad ethical literacy and established issues such as algorithmic bias, fairness, privacy, and social responsibility. In contrast, explainability, trust, and governance receive comparatively less attention in AI ethics curricula, despite their prominence in responsible AI discourse (Reuel, 2025). Explainability is often linked to system trustworthiness, and trust is commonly discussed as a condition for public acceptance and adoption of AI systems (Afroogh et al., 2024)Governance and regulatory frameworks were also underrepresented. Although governance features prominently in policy and ethics discourse (Jobin et al., 2019), limited curricular coverage may leave students less prepared to engage with legislation and policy as they move into professional or civic roles, One study made understanding frameworks and governance an explicit learning outcome, giving students tools to navigate policy expectations (Alam, 2023). Such designs show that governance can be taught as applied practice rather than background context.

Coordination between education and policymaking may strengthen alignment between competency expectations and program design by ensuring that what students are taught reflects evolving governance needs and professional responsibilities. In practice, this could involve shared competency frameworks that translate regulatory and ethical expectations into teachable and assessable learning outcomes, as well as mechanisms for updating curricula as policy guidance changes. A Canadian report on ethical tech innovation recommends collaboration among academia, government, and accrediting bodies (Love et al., 2021), which could support the development of common expectations for AI ethics competencies across programs. Such coordination may also improve comparability

across courses and institutions by encouraging consistent reporting of learning outcomes and assessment approaches, and by reducing disciplinary silos that limit students' exposure to governance, oversight, and policy-relevant dimensions of responsible AI.

Future research should test whether adding modules on explainability, trust, and governance improves performance on scenario-based tasks and policy interpretation. Comparative studies can evaluate curricula with and without these components using shared outcomes and rubrics. Longitudinal designs should assess whether gains persist beyond the term, and interdisciplinary cohorts should examine how preparation differs across technical and non-technical programs.

3.3. Educational strategies

AI ethics instruction employed a wide range of strategies, from lectures and case studies to small-group discussions, role-playing, and applied activities. Lectures were the most frequently reported method, yet they rarely appeared in isolation. Authors typically combined them with active formats such as case-based learning or hands-on projects. This pattern suggested that educators already used blended approaches to engage students in ethical reasoning. Many of the instructional methods observed, such as real-world examples, role-play, and group work, align with established principles for ethics instruction that emphasize dialogical, participatory, and applied learning experiences (Bebeau, 2002; Hartman & Hartman, 2004). However, the wide variation in pedagogical strategy is not matched by an equally robust body of evaluative research.

3.4. Evaluation of learning outcomes

Across the included studies, self-report methods were the most common approach to evaluating learning outcomes, especially surveys and student reflections. While these methods offer valuable insights into students' perceived learning, engagement, and attitudinal shifts, they also present notable limitations. Self-assessment can be useful when measuring changes in attitudes, perceived competence, or learning gains, but its effectiveness varies depending on the context and the design of the tool, and making comparisons across studies or student populations can be challenging if the self-assessment instruments are not validated or standardized (Davis et al., 2023). Thus, it can be difficult to accurately assess actual improvements in ethical reasoning or decision-making skills using self-report alone.

Performance-based assessments appeared less frequently. When evaluation relied on course grades, ethics-specific learning criteria were not always clearly separable from other grading components in the reporting, which may limit interpretability. Digital humanities scholars have critiqued a broader drift toward "ethical compliance," where ethics is treated as rule-following and thus easily graded, which can displace reflective, long-term internalization of ethical thinking (Proferes, 2021). This critique raises concerns about the suitability of grades as a primary indicator of ethics learning in AI contexts.

A major concern identified in this review is the prevalence of purported learning outcomes: claims about student learning outcomes that are presented without accompanying empirical evidence. Many papers outline expected outcomes, such as improved ethical awareness or critical thinking, but do not formally evaluate whether these outcomes are achieved. For example, Alam (2023) describes a course stating that students will "understand the technical capabilities and limitations of AI systems, as well as the potential impact of AI on society, including issues related to safety, fairness, privacy, and ethics". However, the paper does not provide any empirical evaluation to support the claim that students actually learned or will learn this. This pattern was found across multiple studies, where positive outcomes are purported but not substantiated through structured assessment.

A central issue was the lack of empirical assessment across methods. Few studies tested whether specific educational strategies produced measurable gains in ethical understanding, reasoning, or behavior. As a result, it remains unclear which approaches are most effective for different learning goals or student populations. A systematic review that examined empirically reported studies on AI ethics education and interventions across K-12 and post-secondary settings reached similar conclusions (Wiese et al., 2025). It highlighted the need to further develop interdisciplinary AI ethics and to separate educational evaluation from research approaches so that claims about effectiveness rest on sound evidence.

Reporting detail about instruments, scoring, and analytic procedures was often limited, which constrains cross-study comparability. Few papers described how they analyzed results, which limited assessment of rigor and reliability and made comparisons across interventions difficult. More structured and transparent assessment frameworks are needed to move beyond self-report toward evidence-based claims. One option is to anchor outcomes in Bloom's taxonomy, which distinguishes knowledge types and cognitive processes and can guide the selection of aligned measures (Krathwohl, 2002). A taxonomy-based approach makes the link explicit between activities, targeted outcomes, instruments, scoring, and analysis.

Based on the evaluation patterns synthesized in this review and recurring limitations in reporting and analytic transparency, we propose a set of assessment recommendations that may improve comparability across AI ethics courses and strengthen the interpretability of learning claims. These recommendations are offered as practical guidance for aligning outcomes, measures, and reporting, rather than as claims about the effectiveness of any single pedagogy.

1. Learning outcomes should be stated in observable terms that match the claim being made, for example, knowledge, skills, attitudes, or applied judgment. In practice, this means avoiding broad outcomes such as "students will understand AI ethics" and instead specifying what learners will be able to do or demonstrate, such as identifying ethically salient features of an AI deployment scenario, articulating relevant stakeholders and potential harms, or applying a structured framework to justify a course of action. This is consistent with backward-design guidance that treats actionable outcomes as the basis for identifying appropriate "evidence" of learning (Tsunami et al., 2024). This level of specificity helps readers interpret whether the chosen measures plausibly

align with the stated outcomes and reduces ambiguity in claims about learning by making the intended evidence and its relationship to the outcome more explicit.

2. When applied ethical reasoning or decision-making is an intended outcome, include at least one performance-based assessment in addition to self-report measures, because competence-like outcomes can look very different when measured by what students say versus what they do on an applied task (Davis et al., 2023). A feasible minimum is a scenario-based prompt in which students must identify ethically salient issues, weigh competing values, justify a recommendation, and acknowledge uncertainty or trade-offs, scored with an explicit rubric.
3. When surveys are used, prioritize instruments with published validity evidence and documented measurement properties that fit the construct and context, since even widely used higher-education questionnaires can rest on weak-to-moderate and incomplete validity evidence, making uncritical score interpretation risky (O'Neill et al., 2023). Where study-specific items are necessary, report item origins, such as those derived from stated learning outcomes or a conceptual framework, response scales, and key limitations affecting interpretation or comparability. Even brief validity-related reporting would strengthen the credibility of conclusions drawn from survey results and support more meaningful cross-study comparison.

3.5. Reported learning outcomes

Across included studies, authors most commonly reported student's gains in ethical awareness and their ability to recognize issues. It remained unclear, however, whether this awareness translated into deeper ethical reasoning or improved decision-making. Many papers reported heightened awareness or improved conceptual grasp of fairness and bias, but are missing whether students applied principles in practice. This gap points to a need to assess not only what students know, but how they use that knowledge in real or simulated contexts.

The findings raise a curricular question about whether awareness-focused instruction provides sufficient preparation for professional contexts where practitioners must apply ethical principles to authentic challenges. Future work should test designs that deliberately build moral reasoning skills, make ethical commitments explicit, and require students to weigh competing values in ambiguous or high-stakes scenarios. Studies should compare alternative designs using shared outcomes and analytic rubrics, incorporate performance-based tasks and reflective components, follow students longitudinally to assess durability, and include interdisciplinary cohorts to evaluate generalizability.

3.6. Direction of change and long-term impact

Across studies that reported outcomes, authors commonly described positive effects on ethical awareness and thinking. Limited reporting of neutral or negative findings raises the possibility of publication or reporting bias, which can distort judgments about which strategies are genuinely effective. In addition, many papers framed success in broad terms without testing the magnitude of change or the conditions under which it occurred. This pattern limits interpretability because outcomes are often reported as improvement without sufficient detail to assess degree, variation, or uncertainty. More granular evaluations are therefore needed, including attention to differences by student population, instructional strategy, and learning objectives. One practical option for reducing ambiguity is to specify intended learning using a taxonomy-based approach, such as Bloom's taxonomy, which can support clearer alignment between outcome claims and assessment methods by distinguishing cognitive processes (Krathwohl, 2002).

A further limitation is that the literature provides little insight into long-term impact. Short-term assessments dominated the corpus, and we did not identify studies that followed students over time to examine whether ethical understanding is retained or whether reasoning transfers to later academic or professional contexts. This gap is consequential because AI is increasingly embedded in professional and civic settings where ethical judgment must be applied under constraints, including organizational incentives, limited information, and competing stakeholder demands. If the aim of AI ethics education includes preparing learners to navigate such conditions, then evidence is needed on whether learning persists beyond the course and supports application in realistic contexts. Relatedly, the "alignment problem" is often framed as aligning AI systems with human values (Ngo et al., 2022). A complementary challenge is aligning ethics education with the long-term goal of cultivating durable ethical reasoning that remains usable when learners face uncertainty, trade-offs, and institutional pressures. From this perspective, instruction should build immediate understanding while also supporting the development of ongoing reflective capacities, judgment, and responsible action.

To strengthen the evidence base, future research should expand reporting beyond positive effects by including null and negative findings and by specifying the magnitude and distribution of change where quantitative designs are used. Where feasible, studies can preregister outcomes and analysis plans to reduce analytic flexibility and improve interpretability. Designs should also operationalize learning outcomes in ways that permit comparison across contexts, including taxonomy-aligned measures and rubric-scored performance tasks when applied judgment is an intended outcome. Finally, longitudinal designs are needed to examine durability and transfer, including follow-up measures or application-focused assessments linked to internships, capstones, or early professional roles. Comparative designs and interdisciplinary cohorts would further help clarify which instructional approaches support sustained ethical decision-making beyond the classroom.

3.7. Population and disciplinary focus

The disciplinary concentration observed in this review, with most studies focused on computer science and engineering students, suggests that AI ethics education remains framed primarily as a technical concern. This framing prioritizes issues tied to development

and deployment while giving less attention to ethical challenges of use, oversight, and regulation in fields such as law, business, healthcare, design, and public policy. It also contributes to ethical "silos" within AI education, as Javed et al. (2023) note, curricula are frequently compartmentalized by discipline and topic, which limits opportunities for interdisciplinary integration. As AI shapes diverse sectors, a narrow disciplinary focus risks excluding the insights, priorities, and lived experiences of professionals who implement, regulate, and respond to the societal impacts of AI (Javed et al., 2023).

Broadening access to AI ethics education across disciplines is therefore essential. Doing so can cultivate more context-sensitive reasoning and build the cross-sector expertise needed for collaborative and inclusive AI governance. Future work should test interdisciplinary models such as co-taught courses, mixed cohorts, and problem-based projects using shared outcomes and analytic rubrics. Studies should compare results across technical and non-technical programs, include longitudinal follow-ups to assess durability, and examine whether interdisciplinary designs reduce siloing while preserving conceptual depth and technical relevance.

3.8. Potential limitations

We used a Continuous Active Learning (CAL) system to prioritize title and abstract screening. The initial search retrieved 50 766 records. Following developer guidance, screening paused when a screener encountered five consecutive batches of 100 records each with an inclusion rate below 5 percent. Using this stopping rule, we screened only 9 441 records rather than the full set. Because CAL continuously ranks records by predicted relevance, some relevant items can be pushed late in the queue and never reviewed once the yield threshold is reached. As a result, it is possible that eligible studies were missed due to the prioritization and stopping mechanism inherent to the CAL process.

Screening occurred in English only due to team capacity. Relevant studies published in other languages may have been missed, which could bias the corpus toward English-speaking regions and venues. Searches covered 2011 to July 2024. Earlier work and very recent publications outside this window were not assessed. Given the fast pace of AI pedagogy, findings may underrepresent the most recent instructional innovations. Although we searched multiple academic databases, we did not systematically retrieve grey literature such as course websites, syllabi repositories, or internal institutional reports. Excluding these sources may omit applied curricula and evaluations that are not formally published.

We categorized studies as primary, component, or peripheral to focus the synthesis. These labels depend on author reporting and reviewer judgment. Misclassification is possible where papers use overlapping terms for AI, data science, or computing ethics, which could shift counts at the margins. Included studies varied in outcome definitions, instruments, and analytic approaches. This heterogeneity precludes effect-size estimation or formal cross-study comparisons and limits the strength of inferences about relative effectiveness.Most included studies focused on university contexts with undergraduate and graduate students. This skew may overlook insights from polytechnics, colleges, and other post-secondary institutes, limiting what can be inferred about AI ethics education outside university settings.

4. Concluding Remarks

As authors, we interpret the evidence presented in this review as a detailed portrayal of a field in rapid transformation. To foreground the contribution, we distill the synthesis into three core implications for instructional design and future research.

First, course positioning matters. Across the corpus, AI ethics instruction is most often embedded within technical courses (Section 3.3; Table 5), yet the literature lacks comparison of embedded and standalone formats using shared outcomes or analytic rubrics. This gap matters because widely used curriculum and accreditation guidance frames ethical and professional responsibility as a core learning outcome for computing and engineering graduates, making it important to build an evidence base about how different instructional formats support that outcome (Joint Task Force on Computer Science Curricula, 2024; ABET, 2024).

Second, curricula tend to prioritize "teachable" topics, while governance-related competencies lag. Bias, fairness, and privacy dominate curricular attention in the reviewed studies, whereas governance, regulation, and explainability appear less frequently. This imbalance is notable in light of governance-oriented frameworks and emerging regulations that foreground accountability, transparency, and explainability as central expectations for trustworthy AI practice (NIST, 2023).

Third, the synthesis indicates that the field would benefit from shared assessment frameworks that enable a shift from justification to verification in claims about instructional impact. Evaluation practices in the included literature relied heavily on self-report measures and reflective artifacts, and many studies reported positive outcomes without providing sufficient detail to evaluate measurement validity, scoring procedures, or analytic rigor. Related reviews of ethics-education assessment similarly emphasize the limited use of valid and reliable instruments and the need for further instrument development and validation (Kim & Bairaktarova, 2023). As a result, it remains difficult to determine which instructional strategies support particular forms of learning, and whether reported gains extend beyond short-term awareness to more durable reasoning or applied judgment. For future research, the implication is that advances in pedagogy should be accompanied by advances in evaluation infrastructure, including clearer outcome specification, more transparent assessment reporting, and designs that support replication and cross-study comparison.

4.1. Pedagogical interpretation

The results in Section 3.3 and Table 6 show that ethics instruction in AI is most often embedded within technical courses. This model has the practical benefit of accessibility and curricular feasibility but also positions ethics as a secondary concern. Embedding provides context but can restrict the time and conceptual depth needed for genuine ethical deliberation. Standalone courses, which are

fewer in number, demonstrate richer opportunities for philosophical engagement, interdisciplinary dialogue, and reflection. We believe that future pedagogy should combine both approaches, embedding ethics within technical work while also creating dedicated spaces for theoretical and moral reasoning.

The thematic distribution in Section 3.4 and Table 7 show that bias, fairness, and privacy dominate curriculum content, while governance, regulation, and explainability appear less frequently. We interpret this as a reflection of pedagogical pragmatism. Topics such as bias and fairness are concrete, supported by abundant examples and open datasets, which make them easier to teach and evaluate. In contrast, governance and regulatory literacy require interdisciplinary coordination and familiarity with policy frameworks that few technical instructors possess. This imbalance highlights the need to invest in educator training and institutional partnerships that enable teachers to address both individual and systemic dimensions of AI ethics.

4.2. Interpretation of teaching and learning practices

The pedagogical methods presented in Section 3.5 and Table 8 reveal a strong reliance on lecture and case discussion. These methods are effective for foundational literacy but inadequate for developing ethical agency. We understand this as a structural challenge that reflects the composition of teaching teams rather than a lack of pedagogical imagination. Most instructors come from technical disciplines and are more comfortable delivering content than facilitating ethical dialogue or open reflection. Institutional incentives and resource limitations also make active learning more difficult to implement.

The dominance of self-reported evaluation methods described in Section 3.6 and Table 9 further demonstrates the need for validated assessment tools. Surveys and reflective essays provide insight into student attitudes but not into their ability to reason ethically in practice. The field requires standardized instruments and shared rubrics that capture multiple dimensions of ethical learning, including analytical reasoning, empathy, and decision-making under uncertainty. To make this need actionable and comparable across studies, we summarize a minimum set of assessment recommendations in that specifies outcomes, performance tasks, scoring transparency, and durability checks.

The learning outcomes summarized in Section 3.7 and Table 10 show that most courses succeed in raising awareness but fall short of demonstrating behavioural change. This trend is consistent with the early stages of curricular innovation. Awareness and reflection are necessary first steps, but they must evolve into competence and accountability. We believe that longitudinal curriculum design can help link awareness to application by creating continuous opportunities for ethical engagement throughout the degree program.

4.3. Disciplinary and institutional implications

Section 3.8 shows that most AI ethics initiatives are housed within computing and engineering programs. This distribution reflects where public scrutiny and regulatory pressure are most intense. However, it also reinforces a narrow conception of ethics as a matter of technical compliance rather than human judgment. We believe that real progress will depend on interdisciplinary integration. Courses in law, philosophy, social sciences, business, and design should be co-developed with computing departments to reflect the socio-technical nature of AI systems. Institutions should encourage co-teaching arrangements and shared learning outcomes that make ethical reasoning a collective academic responsibility rather than a specialized elective.

The emphasis on undergraduate education also suggests that many initiatives are reactive to institutional mandates or accreditation requirements. Graduate and professional programs remain comparatively underdeveloped. Yet it is precisely in these advanced stages of training that ethical reasoning must mature into professional accountability. Expanding the reach of AI ethics education beyond introductory courses will ensure continuity and depth in ethical development.

4.3.1. Implications for educators and instructional design

For educators designing AI ethics instruction in higher education, the synthesis suggests three practical priorities. First, because embedded delivery dominates the literature, instructors integrating ethics into technical courses should make ethics outcomes explicit and protect time for structured ethical reasoning rather than treating ethics as an add-on. Second, because curricular coverage clusters around bias/fairness and privacy, courses should consider adding structured exposure to governance and regulation, transparency and explainability, and trust and oversight to better reflect the breadth of responsible AI concerns. Third, because evaluation is often based on self-report, instructors can strengthen evidentiary claims by pairing self-report with at least one performance-based activity (for example, a scenario-based task scored with a rubric) that captures applied ethical judgment.

4.3.2. Implications for policymakers and accrediting bodies

For policymakers, professional bodies, and accrediting organizations seeking to strengthen responsible AI capacity, the findings indicate that clearer competency expectations and reporting norms would improve comparability across educational initiatives. Field-level guidance that specifies minimum competencies (for example, ethical issue recognition, applied reasoning, and governance literacy) and encourages transparent evaluation reporting (instruments, scoring procedures, and analytic approach) could help shift the evidence base from descriptive accounts toward cumulative knowledge. In addition, incentives that support interdisciplinary teaching and shared assessment infrastructure may reduce disciplinary siloing and strengthen alignment between educational practice and emerging governance expectations.

4.4. Methodological and theoretical implications

The studies reviewed in Section 3 underline that pedagogical creativity has not yet been matched by theoretical coherence. Many instructors introduce innovative learning activities but seldom relate them to established theories of moral or experiential learning. As authors, we consider this a missed opportunity. Integrating frameworks from education and moral psychology could help explain how ethical understanding is acquired, practiced, and retained. Building on reflective and transformative learning theories can also clarify how students internalize ethical principles and apply them under real constraints. Aligning pedagogy with theory will strengthen the epistemic foundations of AI ethics education and elevate it from an experimental phase to a research-based field.

4.5. Limitations and future directions

We acknowledge several limitations that parallel those observed in the reviewed literature and detailed in The focus on English-language and peer-reviewed publications may have excluded innovative practices from other linguistic or informal educational contexts. Many studies provided limited methodological detail, which restricted our ability to evaluate the quality of evidence. Because this is a scoping review, we did not conduct a formal critical appraisal or statistical power assessment across included studies. These limitations reflect broader structural issues in how AI ethics education research is conducted and reported.

Looking ahead, we believe that future research should prioritize longitudinal and comparative designs that assess how ethical learning is retained and applied beyond the classroom. Studies should follow students into professional environments to examine how ethical reasoning influences real decision-making. Comparative analyses across institutions and teaching methods can help identify which pedagogical strategies are most effective for specific learning outcomes. Collaboration among educators, evaluators, and policymakers will be crucial for building a more cumulative and reliable evidence base.

Reflecting on these findings, we believe that AI ethics education has reached an important threshold. The field has succeeded in establishing visibility and legitimacy across higher education, but it now faces the challenge of depth, coherence, and sustainability. Our synthesis of the literature shows that awareness-raising initiatives have paved the way for a more deliberate phase of curriculum design, one that requires stronger theoretical grounding and empirical validation.

We see this as a pivotal moment for both research and practice. The potential of AI ethics education lies in its ability to move beyond declarative knowledge toward genuine ethical competence. Achieving this will require interdisciplinary collaboration, shared standards for assessment, and institutional support that recognizes ethical reasoning as a professional skill. By coordinating efforts to identify field-level research priorities, there is greater potential for collective societal impact. If pursued with rigor and inclusivity, AI ethics education can become not only a safeguard against technological harm but also a catalyst for cultivating reflective, responsible, and socially engaged practitioners of AI.

Statement of funding

Funding: The Natural Sciences and Engineering Research Council of Canada (NSERC) provided student scholarship funding through Grant ID 554764-2021. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. https://www.nserc-crsng.gc.ca/ase-oro/Details-Detailles_eng.asp?id=722656. We would like to clarify that funding was not granted for this specific project and it has not been peer reviewed. NSERC-CREATE funding was solely for student fellowships. For these reasons, no funding letter is attached.

CRediT authorship contribution statement

Calvin Hillis: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Maushumi Bhattacharjee:** Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **Batool AlMousawi:** Writing – review & editing, Writing – original draft, Validation, Data curation. **Riley Martens:** Writing – original draft, Validation, Data curation. **Tarik Eltanahy:** Validation, Data curation. **Sara Ono:** Validation, Data curation. **Marcus Hui:** Validation, Data curation. **Ba' Pham:** Methodology, Conceptualization. **Michelle Swab:** Writing – review & editing, Methodology, Conceptualization. **Gordon V. Cormack:** Software, Conceptualization. **Maura R. Grossman:** Software, Funding acquisition, Conceptualization. **Ebrahim Bagheri:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization. **Zack Marshall:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ipm.2026.104767.

Data availability

Data is available in supplementary materials

References

ABET Engineering Accreditation Commission. (2024). *Criteria for accrediting engineering programs, 2025–2026*. ABET.

Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: Progress, challenges, and future directions. *Humanities and Social Sciences Communications, 11*(1). https://doi.org/10.1057/s41599-024-04044-8

Alam, A. (2023). Developing a curriculum for ethical and responsible AI: A university course on safety, fairness, privacy, and ethics to prepare next generation of AI professionals. In G. Rajakumar, K.-L. Du, & Á. Rocha (Eds.), *Intelligent communication technologies and virtual mobile networks: 171*. *Intelligent communication technologies and virtual mobile networks* (pp. 879–894). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-1767-9_64.

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Bebeau, M. J. (2002). The defining Issues Test and the four component model: Contributions to professional education. *Journal of Moral Education, 31*(3), 271–295. https://doi.org/10.1080/0305724022000008115

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

Braz, A. P., & Martín del Pozo, M. (2025). Serious games for english language learning: A scoping review. *International Journal of Serious Games, 12*(3), 99–127. https://doi.org/10.17083/t1g08720

Brown, M., Ross, T., Leo, J., Buliung, R., Shirazipour, C. H., Latimer-Cheung, A. E., & Arbour-Nicitopoulos, K. P. (2021). A scoping review of evidence-informed recommendations for designing inclusive playgrounds. *Frontiers in Rehabilitation Sciences, 2*, Article 664595. https://doi.org/10.3389/fresc.2021.664595

Brown, N., Xie, B., Sarder, E., Fiesler, C., & Wiese, E. S. (2024). Teaching ethics in computing: A systematic literature review of ACM computer science education publications. *ACM Transactions on Computing Education, 24*(1), Article 6. https://doi.org/10.1145/3634685. Article.

Buruk, B., Ekmekci, P. E., & Arda, B. (2020). A critical perspective on guidelines for responsible and trustworthy artificial intelligence. *Medicine, Health Care and Philosophy, 23*(3), 387–399. https://doi.org/10.1007/s11019-020-09948-1

Chan, P. Z., Jin, E., Jansson, M., & Chew, H. S. J. (2024). AI-based noninvasive blood glucose monitoring: Scoping review. *Journal of Medical Internet Research, 26*, Article e58892. https://doi.org/10.2196/58892

Cheng, L., Varshney, K.R., & Liu, H. (2021). *Socially responsible AI algorithms: Issues, purposes, and challenges*. https://doi.org/10.48550/ARXIV.2101.02032.

Coldwell, D. A., Venter, R., & Nkomo, E. (2020). Developing ethical managers for future business roles: A qualitative study of the efficacy of "stand-alone" and "embedded" university "Ethics" courses. *Journal of International Education in Business, 13*(2), 145–162. https://doi.org/10.1108/jieb-08-2019-0040

Cormack, G. V., & Grossman, M. R. (2016). Scalability of continuous active learning for reliable high-recall text classification. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (pp. 1039–1048). https://doi.org/10.1145/2983323.2983776

Davis, K. A., Grote, D., Mahmoudi, H., Perry, L., Ghaffarzadegan, N., Grohs, J., Hosseinichimeh, N., Knight, D. B., & Triantis, K. (2023). Comparing self-report assessments and scenario-based assessments of systems thinking competence. *Journal of Science Education and Technology, 32*(6), 793–813. https://doi.org/10.1007/s10956-023-10027-2

Dignum, V. (2017). Responsible artificial intelligence: Designing AI for human values. *ITU Journal: ICT Discoveries, 1*. https://www.itu.int/en/journal/001/Documents/itu2017-1.pdf.

European Commission, High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein center research publication No. 2020-1. https://doi.org/10.2139/ssrn.3518482.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Francombe, J., Ali, G.-C., Gloinson, E. R., Feijao, C., Morley, K. I., Gunashekar, S., & de Carvalho Gomes, H. (2022). Assessing the implementation of digital innovations in response to the COVID-19 pandemic to address key public health functions: Scoping review of academic and nonacademic literature. *JMIR Public Health and Surveillance, 8*(7), Article e34605. https://doi.org/10.2196/34605

Grosz, B. J., Grant, D. G., Vredenburgh, K., Behrends, J., Hu, L., Simmons, A., & Waldo, J. (2019). Embedded EthiCS: Integrating ethics across CS education. *Communications of the ACM, 62*(8), 54–61. https://doi.org/10.1145/3330794

Guizzardi, R., Amaral, G., Guizzardi, G., & Mylopoulos, J. (2023). An ontology-based approach to engineering ethicality requirements. *Software and Systems Modeling, 22*(6), 1897–1923. https://doi.org/10.1007/s10270-023-01115-3

Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2125–2126). https://doi.org/10.1145/2939672.2945386

Hanif, A., Zhang, X., & Wood, S. (2021). A survey on explainable artificial intelligence techniques and challenges. In *2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW)* (pp. 81–89). https://doi.org/10.1109/EDOCW52865.2021.00036

Hartman, L. P., & Hartman, E. M. (2004). How to teach ethics: Assumptions and arguments. *Journal of Business Ethics Education, 1*(2), 165–212. https://doi.org/10.5840/jbee20041212

Hillis, C., Bhattacharjee, M., AlMousawi, B., Eltanahy, T., Ono, S., Hui, M., Pham, B., Swab, M., Cormack, G. V., Grossman, M. R., Bagheri, E., & Marshall, Z. (2025). Teaching postsecondary students about the ethics of artificial intelligence: A scoping review protocol. *PLOS One, 20*(7), Article e0329020. https://doi.org/10.1371/journal.pone.0329020

Javed, R. T., Nasir, O., Borit, M., Vanhée, L., Zea, E., Gupta, S., Vinuesa, R., & Qadir, J. (2023). Get out of the BAG! silos in AI Ethics Education: Unsupervised topic modeling analysis of global AI curricula (extended abstract). *Article, 780*. https://doi.org/10.24963/ijcai.2023/780

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Joint Task Force on Computer Science Curricula. (2024). *Computer science curricula 2023: Curriculum guidelines for undergraduate degree programs in computer science*. Association for Computing Machinery, IEEE Computer Society, & Association for the Advancement of Artificial Intelligence.

Kim, D., & Bairaktarova, D. (2023, June). *Assessment Instruments for Engineering Ethics Education: A Review and Opportunities , Baltimore , Maryland*. https://doi.org/10.18260/1-2–42322

Kopec, M., Magnani, M., Ricks, V., Torosyan, R., Basl, J., Miklaucic, N., Muzny, F., Sandler, R., Wilson, C., Wisniewski-Jensen, A., Lundgren, C., Baylon, R., Mills, K., & Wells, M. (2023). The effectiveness of embedded values analysis modules in computer science education: An empirical study. *Big Data & Society, 10*(1). https://doi.org/10.1177/20539517231176230

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Lefebvre, C., Glanville, J., Briscoe, S., Featherstone, R., Littlewood, A., Metzendorf, M.-I., Noel-Storr, A., Paynter, R., Rader, T., Thomas, J., Wieland, L. S, et al. (2025). *Chapter 4: Searching for and selecting studies* (Version 6.5.1, last updated March 2025). In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, et al. (Eds.), *Cochrane handbook for systematic reviews of interventions (Version 6.5.1)*. Cochrane. https://www.cochrane.org/handbook.

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science : IS, 5*, 69. https://doi.org/10.1186/1748-5908-5-69

Love, H., Lajoie, J., & Boger, J. (2021). *Ethical tech innovation: Uniting educational initiatives and professional practice*. University of Waterloo. https://dspacemainprd01.lib.uwaterloo.ca/server/api/core/bitstreams/7da22305-f1c5-4a79-9dd4-4ff8c49235e2/content.

Mevlevioğlu, D., Tabirca, S., & Murphy, D. (2023). Anxiety classification in virtual reality using biosensors: A mini scoping review. *Plos One, 18*(7), Article e0287984. https://doi.org/10.1371/journal.pone.0287984

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics, 26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI rmf 1.0)*. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.

Ngo, R., Chan, L., & Mindermann, S. (2022). *The alignment problem from a deep learning perspective* (Version 8). *ArXiv*. https://doi.org/10.48550/ARXIV.2209.00626

OECD. (2019). Recommendation of the council on artificial intelligence. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

O'Neill, L., Lauridsen, H. H., Østengaard, L., & Qvortrup, A. (2023). Validity evidence for the experiences of teaching and learning questionnaire (ETLQ) in evaluations of quality learning: A systematic critical literature review. *Studies in Educational Evaluation, 78*, Article 101283. https://doi.org/10.1016/j.stueduc.2023.101283

Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis, 18*(10), 2119–2126. https://doi.org/10.11124/JBIES-20-00167

Proferes, N. (2021). What ethics can offer the digital humanities and What the digital humanities can offer ethics. In K. Schuster, & S. E. Dunn (Eds.), *Routledge international handbook of research methods in digital humanities* (pp. 416–427). Routledge, Taylor & Francis. https://doi.org/10.4324/9780429777028.

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

Reuel, A. (2025). *Artificial intelligence index report 2025: Chapter 3: Responsible ai [Report]*. Stanford Institute for Human-Centered Artificial Intelligence. https://hai.stanford.edu/assets/files/hai_ai-index-report-2025_chapter3_final.pdf.

Richardson, B., & Gilbert, J.E. (2021). *A framework for fairness: A systematic review of existing fair AI solutions*. https://doi.org/10.48550/ARXIV.2112.05700.

Sarraf-Yazdi, S., Teo, Y. N., How, A. E. H., Teo, Y. H., Goh, S., Kow, C. S., … Krishna, L. K. R. (2021). A scoping review of professional identity formation in undergraduate medical education. *Journal of Gtriccoeneral Internal Medicine, 36*, 3511–3521. https://doi.org/10.1007/s11606-021-07024-9

Shen, L., Chen, I., Grey, A., & Su, A. (2021). Teaching and learning with artificial intelligence. In S. Verma, & P. Tomar (Eds.), *Advances in educational technologies and instructional design* (pp. 73–98). IGI Global. https://doi.org/10.4018/978-1-7998-4763-2.ch005.

Sowmia, K. R., & Poonkuzhali, S. (2020). Artificial intelligence in the field of education: A systematic study of artificial intelligence impact on safe teaching learning process with digital technology. *Journal of Green Engineering, 10*(4), 1566–1583. Scopus.

The 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. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html.

Tsunami, C. K., Henríquez-Trujillo, A. R., Ferreira-Meyers, K., Mwanda, Z., Rimal, J., Pozu-Franco, J., & Delvaux, T. (2024). Guidelines for integrating actionable A-SMART learning outcomes into the backward design process. *MedEdPublish, 14*, 242. https://doi.org/10.12688/mep.20606.1

Turner Lee, N. (2018). Detecting racial bias in algorithms and machine learning. *Journal of Information, Communication and Ethics in Society, 16*(3), 252–260. https://doi.org/10.1108/JICES-06-2018-0056

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0850

UNESCO. (2021). Recommendation on the ethics of artificial intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000380455.

Venkatasubramanian, S. (2019). Algorithmic fairness: Measures, methods and representations. In *Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems* (p. 481). https://doi.org/10.1145/3294052.3322192

Wiese, L. J., Patil, I., Schiff, D. S., & Magana, A. J. (2025). AI ethics education: A systematic literature review. *Computers and Education: Artificial Intelligence, 8*, Article 100405. https://doi.org/10.1016/j.caeai.2025.100405

Yin, J., Ngiam, K. Y., & Teo, H. H. (2021). Role of artificial intelligence applications in real-life clinical practice: Systematic review. *Journal of Medical Internet Research, 23*(4), Article e25759. https://doi.org/10.2196/25759

Zuber, N., Kacianka, S., & Gogoll, J. (2022). Big data ethics, machine ethics or information ethics?. *Navigating the maze of applied ethics in it (Version 1)*. https://doi.org/10.48550/ARXIV.2203.13494. arXiv.