Computers and Education: Artificial Intelligence 8 (2025) 100405

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

AI ethics education: A systematic literature review

Lucas J. Wiese [a,*], Indira Patil [b], Daniel S. Schiff [b], Alejandra J. Magana [a,c]

[a] Department of Computer and Information Technology, Purdue University, 401 Grant St, West Lafayette, IN, 47907, USA
[b] Department of Political Science, Purdue University, 401 Grant St, West Lafayette, IN, 47907, USA
[c] Department of Engineering Education, Purdue University, 401 Grant St, West Lafayette, IN, 47907, USA

ARTICLE INFO

Keywords:
AI ethics
Systematic literature review
AI education
Ethics education
Assessment
Measurement

ABSTRACT

The potential of AI technology to transform human life, well-being, and daily work is faced with numerous risks and challenges yet to be fully accounted for. However, the complexity of AI ethics makes it hard to pin down *what* to teach, *how* to teach it, and how to *assess* its effectiveness. Drawing on an educational perspective, this paper presents a systematic literature review and qualitative analysis of the early years of AI ethics education as a formalized field to analyze whether its future trajectory is aligned with educational best practices. Our review highlights core challenges in AI ethics education and the content, assessment, and pedagogy used in real interventions over recent years. We find that efforts to teach AI ethics do helpfully draw on a holistic view (as opposed to a narrow view), and utilize progressive pedagogies like case studies and group projects that aim to meaningfully challenge students' ethical reasoning skills in applied practices. However, many real- world AI ethics teaching interventions do not leverage well-supported assessment techniques known to support student learning; rather, assessment is conducted primarily for research evaluative purposes. This gap in rigorous assessment raises implications for researchers and practitioners, as responsible development and use of AI will be stymied if educators cannot successfully determine whether students have truly learned relevant AI ethics content or skills.

1. Introduction

> "Today's students will be tomorrow's algorithm engineers; and, through their everyday practice (whether or not they are aware of it), they will weigh efficiency and profit against public safety, justice, fairness, and good for humanity." (Howley et al., 2022, p. 256).

The scholarly discussion of artificial intelligence (AI) began in the 1950s with attempts to model the human mind borne from mathematical thought experiments for computation and has long been a science fiction fan-favorite trope embodying the progress, peril, and persistence of technological development within society. But in the last decade, AI has taken a more serious part of contemporary society's concern. These conversations take multiple flavors: AI's power to displace workers, exacerbate inequality, harm day-to-day human autonomy, intensify cybercrime, and promote mis/disinformation, among other matters (D. Nguyen & Hekman, 2024). Meanwhile, more technical topics of AI-driven bias, unfairness, inexplicability, and proliferation of misinformation constitute additional concerns (Kazim & Koshiyama, 2021; Kuipers, 2020). These diverse issues understandably have driven repeated calls to educate the workforce, and the public, on AI ethics. Indeed, a heightened level of awareness, knowledge, and skill is required across society to ensure the sustainable development and use of responsible AI.

> Americans of every age, in every district, and from every background will be impacted by AI, and therefore need AI literacy–an understanding of basic AI principles and applications, the skills to recognize when AI is employed, and awareness of its limits. (H.R.6791 - 118th Congress (2023–2024), 2023)

That excerpt, from the bipartisan US bill for the AI Literacy Act, is supported by 28 organizations and signals the critical importance of education for AI. Thi s attention extends internationally, too; As identified by Schiff (2022), 21 of 24 countries' national AI policy strategies included a *dedicated* section on education for AI, and all considered some aspect of training AI experts, workforce preparation, or public literacy. Students themselves wish they were taught more about AI: Workforce surveys report that 70-80% of students feel lackluster AI preparation, and similar figures indicated that employers share this goal (Cengage, 2024; Digital Education Council, 2024). To respond to these national,

* Corresponding author.
E-mail addresses: lwiese@purdue.edu (L.J. Wiese), ipatil@purdue.edu (I. Patil), dschiff@purdue.edu (D.S. Schiff), admagana@purdue.edu (A.J. Magana).

https://doi.org/10.1016/j.caeai.2025.100405
Received 28 October 2024; Received in revised form 6 April 2025; Accepted 6 April 2025
Available online 14 April 2025

institutional, and student-level needs, universities need to prepare for the changing educational landscape in the age of AI.

Numerous regulatory, policy, and educational initiatives (see Arnold et al., 2024) now aim to promote the *responsible* development of AI in particular. For instance, the U.S. Executive Order and the EU AI Act (European Union, 2024; The White House, 2023) represent key attempts to formalize responsible AI guidelines, with both government and non-government organizations pushing change (Fjeld et al., 2020; Jobin et al., 2019). Subsequently, major technology firms like Google, Adobe, IBM, and NVIDIA are investing in workforce and public AI literacy efforts to train students, educators, and workers on AI and responsible AI practices (California, 2024; DeVon, 2024). Many of these initiatives do in fact promote "AI ethics" as a critical part of this digital transformation; however, it remains unclear how to effectively instill this ethical and social lens along with core technological literacy.

Critically, the field of AI ethics is not as simple as teaching future AI experts which ethical principles to abide by. AI ethicists and the many other individuals who need to be literate in AI ethics also need to develop deep critical thinking skills, understand the sociotechnical implications of technology, and practice ethical reasoning that transfers across AI use-cases and contexts. For example, Gambelin (2021) argues that AI ethics competencies include interdisciplinary knowledge of technology, business, law, and policy around AI, on top of communication and leadership skills. Yet, even teaching "ethical reasoning" reliably is a major challenge, as humans have limited cognitive capacity, biased or motivated reasoning, and other unconscious mechanisms at play (Bargh, 2022; Lapsley & Hill, 2008). In short, teaching future practitioners or the broader public to truly understand and practice AI ethics is very difficult, and much is unknown about how AI ethics education is proceeding or should proceed.

In light of the pressing need to educate practitioners and the public about AI ethics, and modest knowledge of how this effort is faring, this paper implements a systematic literature review (SLR) of AI ethics education. It aims to provide a comprehensive picture of the emerging years of AI ethics education as a formalized topic, covering 2018 through early 2023. We analyze real-world interventions in particular (rather than conceptual or abstract research) across dimensions of *content*, *assessment*, and *pedagogy* to examine how complex educational practices are manifesting in practice, rather than just in theory.

While AI ethics education may be relatively new as a formalized field of practice, it is intertwined with prior educational efforts. To ground this analysis, we therefore pay careful attention to and draw on closely related fields: computing ethics education and engineering ethics education. Our focus is on how AI ethics education practices may align, or diverge from trends and best practices in these fields. It is intended to enable examination of the emerging years of AI ethics education and its trajectory as a foundation for future understanding.

As such, we ask the following two primary research questions.

- **RQ1.** What do the early years of (formalized) AI ethics education interventions, from 2018 to 2023, suggest about the state and future trajectory of the field?
- **RQ2.** To what extent does AI ethics education adopt or diverge from best practices in computing and engineering ethics education?

We find that 'progressive approaches' to teaching and pedagogy such as case studies and group projects are quite common in AI ethics education interventions but accompanied by limited educationally effective assessment methods. Further, these interventions emphasize a broader sociotechnical perspective on AI rather than a technical perspective alone. After introducing our conceptual background and methodology in more detail, we describe key descriptive results, thematic results, and close with practical and scholarly recommendations. Overall, the landscape of AI ethics education research has strong and well-conceived roots, but with opportunities for foundational improvement in the future.

2. Background

2.1. The state of AI ethics

Debates around the ethics of AI often invoke discussions about ethical principles, such as transparency, justice, responsibility, and beneficence, among others, which fundamentally help us understand AI's implications in terms of competing human interests (Jobin et al., 2019; Waelen, 2022). Indeed, many of the trends and patterns in responsible AI innovation are driven by a principles-first approach. For instance, major scholarly reviews from Jobin et al. (2019) and Floridi et al. (2018) characterize AI ethics according to key principles like beneficence, non-maleficence, autonomy, justice, and explicability. Following suit, Schiff et al.'s (2020, pp. 153–158) review of 88 AI ethics policies highlights 25 ethics principles across public, private, and non-governmental sectors, and analyzes them to argue how dominant multinational actors motivate and impact the field. Corrêa and colleagues' (2023) meta-analysis reviewed 200 AI policy documents and extracted 17 categories of AI principles, with transparent, reliable, and 'just' AI systems taking the three top spots.

While framing AI ethics issues around these prominent principles can help support an understanding of how AI impacts humans and can provide a conceptual toolkit to shape practice, policy, and AI design (Lu et al., 2024; Umbrello & Van De Poel, 2021), principles are not the end-all-be-all for AI ethics. Indeed, principlism in AI ethics faces various—sometimes strong—critiques which are pertinent to educational design (Floridi, 2019; Hagendorff, 2022a; Munn, 2023). In particular, scholars correspondingly advocate for practitioners to understand not only AI ethics principles or issues but also how to *implement* ethics into the design of AI systems, whether through technological design or governance (Kazim & Koshiyama, 2021). In the development lifecycle of an AI system, sociotechnical considerations can be applied at each phase to translate ethical principles into practices, with implications extending from technical ethics to business ethics and even societal ethics (Martin, 2019).

Proposed solutions are thus both sociotechnical and technical in nature. For instance, a company can conduct participatory stakeholder meetings to promote justice in the design of an AI system (Morley et al., 2021). Or, computer scientists can implement SHAP values to explain how different model features influence the output of an AI system to better design for explicability (Lundberg & Lee, 2017). Developers can systematically document decisions made during the development phase to improve transparency and accountability (Kroll et al., 2017), as well as using computational tools to test the explicability and fairness of models post-deployment (Morley et al., 2020). The depth of transformation needed to apply AI ethics entails that this effort is not a single-person job, nor tractable through a single-use tool. Instead, it requires consistent reflection, communication, and deliberation, with many people thinking across disciplinary lines. In summary, AI ethics is extremely complex; AI ethics education correspondingly must be similarly complex.

2.2. The state of AI education

Scholarship on AI and education has a long history; using AI to improve educational practice was a core interest of the 1956 Dartmouth workshop where "AI" began as a field (Doroudi, 2022). However, it is important to draw a distinction between using AI as a *tool* in education and the distinct body of research on education *for* AI, or "AI literacy" (the focus in the current paper is the latter, emphasizing AI *ethics* literacy in particular). [1] There are three common goals for AI literacy efforts:

[1] Outside the scope of the current study, we recognize the importance of the ethics of *incorporating* AI into education. To start, see Akgun and Greenhow (2022), Dwivedi et al. (2023), and Holmes and Porayska-Pomsta (2022).

training AI experts or specialists, preparing the workforce for digital transformation and improving the public's foundational AI literacy (Schiff, 2022). That is, AI education is not limited to formal education, nor is it targeted at future computer scientists alone; rather it includes public literacy and workforce development efforts (UNESCO, 2022a; 2022b).

AI literacy–which is broader than AI ethics literacy alone–includes the ability of an individual to know and understand AI, apply and use AI, evaluate and create AI, and AI ethics (Ng et al., 2021). More comprehensively, Long and Magerko (2020, pp. 1–16) enumerate a set of 17 competencies for AI literacy ranging from the interdisciplinarity nature of AI to its decision-making capacities, physical action potential, limitations, and more. Tenório and Romeike (2023, pp. 1–12) present a list of 52 AI competencies, focused on non-computer science professionals while nationally-supported initiatives such as 'AI for K-12' (AI4K12, 2020) provide concrete guidelines and age-specific resources for AI education. Private companies have undertaken initiatives emphasizing AI literacy as well. For instance, Google and NVIDIA have committed resources to improve AI literacy, targeting millions of learners (California, 2024; Google, 2024).

Educators looking for help in navigating the technical, ethical, or logical challenges can turn to specific internal institutional guidance, high-level international guidance, or scholarly literature on foundational principles for AI in education (A. Nguyen et al., 2023; UNESCO, 2023; Wiese & Magana, 2024). However, implementing a curriculum that can adequately cover the complex concepts and skills suggested by these frameworks is highly complex, and there is no one-size-fits-all approach. Along these lines, Dai et al. (2023) report on the "negotiation" that takes place when creating and implementing material, such as the need to balance local challenges in students' lives with the external conditions and pressures around them. This complex design exercise persists even when focusing on AI ethics literacy alone, in part because of the numerous subtleties discussed in the prior section.

2.3. What is AI ethics education?

As discussed, AI ethics is a field that aims to help guide the development of AI according to responsible practices to secure and maintain human flourishing and to prevent the impingement of human rights (Hagendorff, 2022b; Lauer, 2021). The field of AI ethics education, then, refers to the teaching and learning of AI ethics. We, therefore, define ***AI ethics education*** *as formal or informal education, of students, the workforce, or public, to promote the responsible development and use of AI, help people understand the benefits and risks of AI, and arm people with the knowledge and skills to navigate contemporary life and work alongside AI.*

Helpfully, many notable AI literacy frameworks do integrate AI ethics, including frameworks by Chiu et al. (2022), Ng et al. (2021), and Southworth et al. (2023). For instance, Chiu and colleagues' (2024) framework explicitly includes both the 'ethics of AI' *and* the 'impact of AI' as distinct aspects of literacy alongside technological knowledge, collaboration skills, and self-reflective capacities. Yet, even when incorporating ethics alongside technical content is widely promoted by ethics education scholars (historically in engineering and computing education and now in AI education) and can indeed lead to better ethical reasoning skills (Grosz et al., 2019), it often falls by the wayside—"if time allows"—in the curriculum (Garrett et al., 2020, pp. 272–278).

Reassuringly, numerous national efforts and university programs are beginning to embed AI ethics across diverse sets of students and citizens. For instance, Finland is undertaking a widespread public AI literacy initiative through the use of massive online open courses (MOOC) in educational and non-education settings (European Commission, 2021). Singapore has similarly supported innovative AI governance, development, and educational practices for the literacy of its citizens with a dedicated AI ethics course (AI.Singapore, 2024). Various universities such as the Massachusetts Institute of Technology (MIT), Stanford University, and Purdue University, have likewise launched responsible AI programs inclusive of educational components. MIT deployed a university-wide initiative for responsible AI (MIT, 2024), Stanford has long honed its Institute for Human-Centered AI, and Purdue carries an MS in AI with required coursework in AI ethics for both STEM and non-STEM students (Purdue, n.d.).

In sum, there is widespread consensus on the importance of AI, AI literacy, and AI ethics education, and some early initiatives. Yet, AI ethics education is bound to transform as both AI and AI ethics evolve (Borenstein & Howard, 2021), and these efforts remain nascent and largely untested, motivating the current review. While some prior reviews have examined AI ethics education before, this scholarship has focused on the ethics of using AI tools *in* education (Dieterle et al., 2022; Holmes & Porayska-Pomsta, 2022), or has covered ethics only as a minor aspect of AI literacy efforts. This paper remediates this gap.

2.4. Leveraging engineering and computing ethics education

This effort follows in the scholarly tradition of computing and engineering education research, which face similar challenges in implementing effective ethics education amidst predominantly technical material. These fields have, at this point, an extensive body of literature that enables even field-specific SLRs (Borrego et al., 2014) including disciplinary reviews in subfields of STEM, computing, and engineering education (Hartikainen et al., 2019; Hess & Fore, 2017; Karabulut-Ilgu et al., 2018; Lyon & Magana, 2020; Van den Beemt et al., 2020). For example, Hess and Fore's (2017) review analyzed engineering ethics interventions and the degree to which ethics goals are implemented, finding overall limited empirical work in engineering ethics education. Martin et al. (2021) scope education at multiple societal levels and analyze morally salient goals or objectives in the engineering education space, situating ethics within broader engineering education paradigms and reform programs. Padiyath (2024) reviews computing courses and examines the educational contexts under which ethics interventions are effective, and which are not, identifying the need to situate ethics curriculum amongst the students as well as broader curricula objectives. Most similarly, Brown et al. (2024) analyzed the conception, delivery, and evaluation of ethics in computing education, finding the need to take ethics measurement and evaluation seriously beyond simple measures.

To connect the present review with this related literature, we draw on the content, assessment, and pedagogy (CAP) framework, which was originally intended to help instructors build effective outcome-based educational practices (Streveler & Smith, 2020) as per the backward design model by Wiggins and McTighe (2005). In this framework, instructors should first start with the end in mind–what learners need to know–then identify the mechanisms that will help the student achieve their learning. Importantly, assessment should play dual roles, providing feedback for learners as well as allowing for summative evaluation and grading. The core emphasis of CAP is that educational programs must be *aligned* across content, assessments, and pedagogy to be effective (Streveler et al., 2012; Streveler & Smith, 2020).

Notably, CAP has been increasingly utilized in studies of computing and engineering ethics education in particular. It serves as a simple but powerful framework that inspires the design of our research questions, codebook, and key elements of our thematic analysis, and which allows us to connect back to insights from computing and engineering ethics education literatures. Along these lines, we next detail the methods chosen to investigate the AI ethics education field in this manner.

3. Methods

3.1. Research overview and rationale

The early years of AI ethics education research (defined from 2018 to early 2023) represent a period of formalization, where the field became more distinct beyond digital, computing, or engineering ethics alone.

While AI ethics education existed tangentially within these domains, the early 2020s marks more concentrated scholarship a paradigm shift before and surrounding the widespread deployment of commercial language models like OpenAI's ChatGPT in late 2022.

While this period is just a snapshot of an evolving topic, reviewing it can highlight the root intentions of AI ethics education researchers and practitioners and suggest a roadmap for where the field may head (including promises and pitfalls). Importantly, however, published accounts of research during this time do not fully capture the range of practices in the field: much of educational practice goes unreported and unpublished, and other methodologies like syllabi analyses and surveys should be utilized in the future.[2] However the chosen methodology allows us to glean insight into how early practices amongst avid and early AI ethics education adopters may direct the future of AI ethics education–or, perhaps provide insufficiently strong foundations for the future of the field.

3.2. Document collection and systematic review process

We adhered to the PRISMA methodology to collect and systematically report document collection and utilized the PRISMA checklist to structure the initial phases of the study procedures: defining research objectives, specifying eligibility criteria, identifying databases and search protocols within different databases, and orderly reporting on database search results (PRISMA, 2020). The following section and Fig. 1 briefly report on our methodological procedure, and Appendix B provides more detailed PRISMA documentation.

In what manner does education about AI incorporate ethics? This question opened our conception of the current study. Here, we are concerned with how education about AI, then, the nature of ethics within these educational interventions about AI. Given the focus on the early years of the field, and our conceptual focus on the content, assessment, and pedagogy pillars of educational design, our research question (RQ) is chiefly about how early years of educational research implicate the trajectory of the field at large. Our primary research question is complemented by an operational question, and a secondary research question approaches broader implications about how AI ethics education is aligned, or unaligned, with best practices in adjacent disciplines of computing and engineering. Thus, our research questions follow.

- (**RQ1**) *What do the early years of (formalized) AI ethics education interventions, from 2018-2023, suggest about the state and future trajectory of the field?*
- (**RQ1.1**) *To what extent are content, assessment, and pedagogy aligned for AI ethics education interventions in the early years of AI ethics education?*
- (**RQ2**) *To what extent does AI ethics education adopt or diverge from best practices in computing and engineering ethics education?*

We conducted our literature search and exported records in January 2023. We used the same search string to query for publications across five databases (with varying syntax, see Appendix B). We initially subsetted to AI ethics research, and then filtered to educational contexts via manual screening, which we found to be a viable strategy for comprehensiveness. Eligible records had to be published between 2018 to January 2023, written in English, and published in conference proceedings or journal venues. Given the novelty of AI-related papers and publications, the research team made note *not* to restrict the search to only top journals or conference venues. The research team identified the following databases for searching records: Web of Science, Scopus, PhilPapers, ACM Digital Library, and IEEE Xplore.[3]

The completed **search export** yielded a total of 2668 records, after which we excluded papers based on additional inclusion and exclusion criteria (Appendix B for full details). **Phase 1** excluded duplicate, non-journal, and non-conference records, and resulted in 1567 records. **Phase 2** screened for 'education' using a set of keywords and settled with 681 records. In **Phase 3**, we appraised each abstract to determine whether keywords were out of context; Sometimes "learning" was "*machine* learning, or "training" was conjoined with "*model* training," and consequently removed 515 records–leaving 166 records. In **Phase 4**, we read the full-text together in team meetings to determine whether the paper was truly an educational intervention on AI ethics: 73 papers were excluded when education was not a substantive focus, 14 were removed when ethics was not within scope, 12 removed due to AI *for* education, and 11 for metadata-related criteria. In the resulting 56 records, we were only concerned about papers with empirical data, concluding with a set of **25 papers** on AI ethics educational interventions.

3.3. Codebook development and codebook procedure

Following our document collection, a deductive coding procedure formed the basis of our thematic analysis by directing our coding procedure from existing theory, research questions, and CAP conceptual framework (Bingham & Witkowsky, 2022). The three pillars of CAP guided the creation of our codebook, which was divided into four sections: metadata and descriptives, ethics content, assessment mechanisms, and pedagogical techniques. The 11 associated code families contained a total of 81 codes, which helped align document collection, coding, analysis, and summarized findings at the end of our research. See an overview of the codebook in Table 1, with the full list of codes and definitions in Appendix A.

3.4. Thematic analysis approach

Thematic analysis is a qualitative approach that identifies, and organizes, trends across various parts of a dataset to form a cohesive picture of the *meaning* in a dataset (Braun & Clarke, 2012) organized into certain themes (Vaismoradi et al., 2016, p. 101). We chose a deductive approach to guide our analysis of RQs by pre-existing theories and concepts. In turn, we were able to transform a large heterogeneous dataset (academic papers) into a structure of codes, and as such, into salient takeaways and themes across our dataset.

Operationally, following Vaismoradi et al.'s (2016) phases of theme development, we first **initialized** the data while three authors coded each paper in ATLAS.ti 23.4.0 (ATLAS.ti, 2023) wrote reflective notes during, and after, coding each paper. We iteratively developed the codebook; Starting from a rough conception of our project and then creating a codebook relevant to our dataset, team norms, and personal strengths and weaknesses through an intercoder agreement (ICA) procedure. We exported each paper's codes to run ICA calculations to ensure an account of the field and give internal metrics for our teams' blind spots. For instance, we initially included a code category for *normative frameworks,* but we realized we could not reliably find and code different frameworks in the paper and removed them from our codebook. After numerous coding and codebook development sessions, adding or removing codes to align with our research objectives, we agreed on a final set of codes. Then, we verified our **constructed** code categories as relevant to answering our research questions by reviewing the codes and memos together. Next, we took a step back from the coding procedure and reviewed our codes, along with our reflective memos, with respect to established field knowledge to **rectify**

[2] For example, one inherent limitation of examining published academic literature on these interventions is that these studies may be skewed toward *novel* pedagogical approaches compared to unpublished interventions.

[3] The research team consulted with an engineering libraries specialist and interdisciplinary subject matter experts during these initial phases of the project.

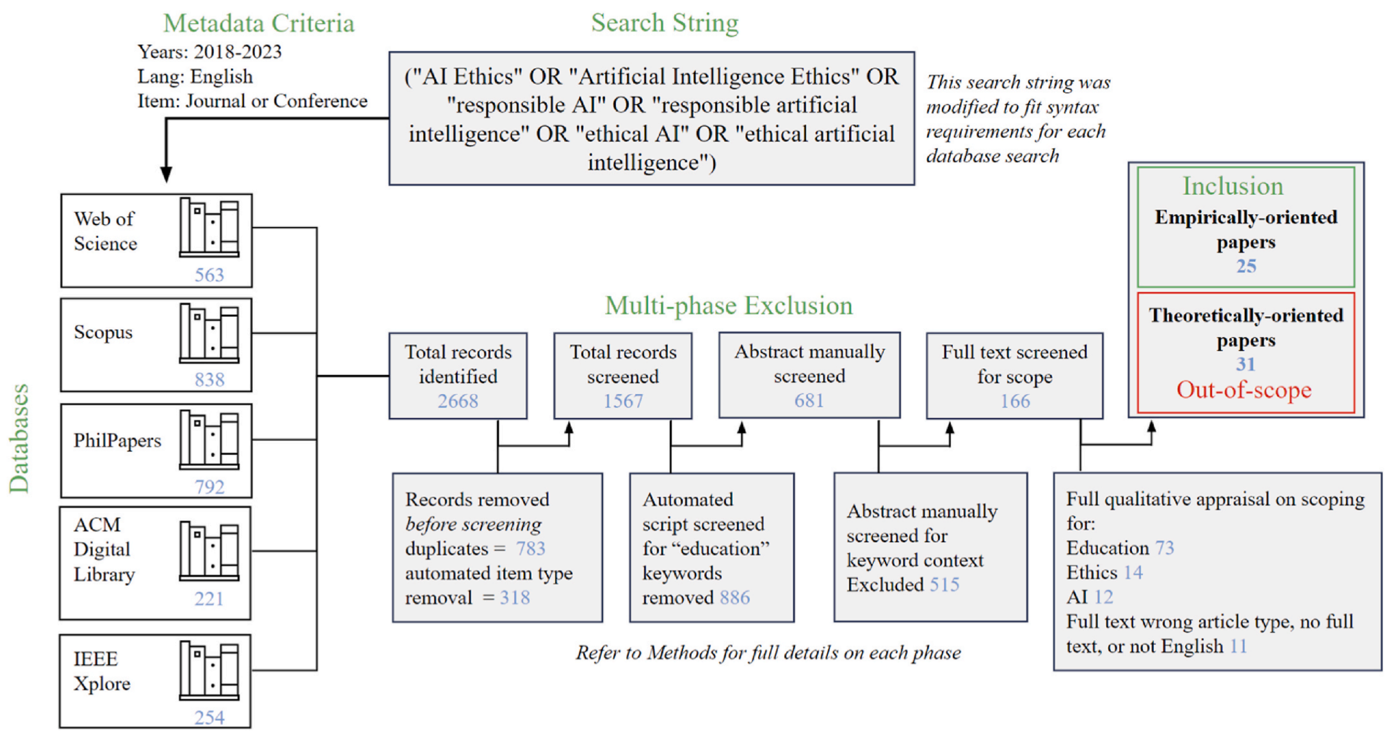

Fig. 1. Methods flow diagram from search string through study identification, screening, and inclusion procedure. Adapted from PRISMA (Page et al., 2021) and Khan et al. (2022, pp. 383–392). Official PRISMA figure in Appendix B.

preliminary key findings for each code family and identify emerging themes in our data. Finally, through meetings and deliberation, we **converged on a final set** of three themes that emerged from our data and were meaningful to understanding the state of AI ethics education research.

4. Results

The following section first provides the descriptive statistics for our papers. Then, core findings associated with each code family are reported in Table 3. We provide a high-level review of the emerging challenges in the data in Table 4 and finally lead into the primary three crosscutting, emergent themes. Correspondingly, **the first theme** suggests that ethics education for AI is responding to the complex nature of the technology by incorporating progressive instructional methods. **The second theme** captures how authors balanced mechanistic and technical instruction of ethics with socially conscious and broad-sweeping societal implications of AI technology. However, these positive trends are met with haphazard student-centered assessment. Thus, **the third theme** analyzes the dynamic between research evaluation and student assessment, finding a lack of formative assessment and little adherence to learning objectives. We hope that more established assessment techniques from the engineering and computing disciplines can inform formative assessment for AI ethics education moving forward.

To answer **RQ1**, we found the early formal years of AI ethics education ambitiously uses advanced pedagogical methods and with societally relevant content matter, but our data suggests challenges implementing practical student-centered assessment. Furthermore, to answer **RQ1.1**, we found that ethics content responds to the complexity of AI itself, and this is met with pedagogical methods that push students to engage in material beyond mere knowledge, but it is likely that students were not given formative feedback during their learning to self-regulate their learning experience. Then, to answer **RQ2**, we find that the field has begun breaking the molds of traditional engineering and computing ethics education. Students are pushed to create things and discuss contentious issues, without dwelling on deep ethical theory and standardized codes. Now, to better understand these questions, we present the following findings with as much clarity and detail as possible.

4.1. Descriptive statistics

The landscape of educational research on AI ethics interventions is predominantly shaped by universities; Of the 25 papers analyzed, 23 papers were produced where the first author's primary affiliation was a university. The two exceptions, Gong et al. (2020) from a research laboratory and Chklovski et al. (2020, pp. 34–35) from an education nonprofit. Moreover, gender diversity is present in the authorship teams, where 23 of the 25 papers had a woman or non-binary author included in the authorship team. International collaboration was moderately present, with 25% of the papers consisting of an international author team. Please see Fig. 2 for the countries represented and the distribution of papers by year published. The interventions took form both within traditional curricula and in extracurricular spaces like workshops and ranged from taking place during one-day to a full academic year. However, none of the interventions took place "across the curriculum" which is frequently recommended in the literature (Grosz et al., 2019; Mitcham & Englehardt, 2019). Please see Table 2 for a breakdown of these intervention strategies and their context.

4.2. Key findings and challenges

To begin the substantive analysis, Table 3 reports on the key findings for each code family. Then, we transition into a thematic exploration of our research questions. In turn, we can provide a snapshot-in-time of educational interventions from 2018 to early 2023. In the following sections, we will further break down and cross-sectionally analyze what these mean as a full picture of AI ethics education.

We will now highlight some of the core challenges of AI ethics education found in our dataset. Table 4 reviews how authors expressed difficulty in navigating educational features, understanding "ethics" as a field, teaching the complexity of AI content, and solving human

Table 1
Codebook by family with description and associated codes. See Appendix A for full Codebook.

Code family	Description	Codes
Metadata	The metadata codes gathered descriptive data about the papers and authors of the papers for **context**.	Gender, international collaboration, quantitative or qualitative methodology, research design, publishing sector
Target population	Codes within this family were adapted from ISCED (Schiff, 2022; UNESCO, 2011). These codes provided **context** on the target population of an educational intervention.	AI experts, business leaders, policymakers, public, STEM workforce, primary students, secondary students, postsecondary students, graduate students, teachers, other
Ethics learning goals	Codes within this family, created by the research team, captured the intended learning outcomes for the ethics **content**.	Communication, ethical reasoning, experience or intuition, knowledge or facts, other
Educational context	Codes within this family were adapted from (Dorie et al., 2012) to capture simple constructs of active and collaborative learning within the educational strategies and interventions. Considered as part of the **pedagogical** environment for these interventions.	active learning, passive learning, collaborative learning, self-directed learning
Ethics topics	These codes were used to capture the topics of the ethics **content** in these interventions. These were adapted from a taxonomy for ethics topics in (Barkhuff et al., 2024).	autonomy, bias, deception, diversity-equity-inclusion, ethical theories, codes of ethics, inequality-power-fairness, intellectual property, mis/disinformation, policing, privacy, risk or safety, robotics, social-good-computing, social justice, sustainability, technological access, transparency or explainability, universal design, other
Instructional delivery type	These codes captured the mechanism of instructional delivery and captured the **pedagogical** environment of these interventions.	Online, in-person, hybrid
Pedagogical techniques	These codes captured the **pedagogy** that authors utilized to deliver their ethics interventions. These were adapted from (Hess & Fore, 2017).	Application of theory-code-law, case study, discussion, experimentation, group project, hands-on learning, lecture, non-traditional, real-world exposure, written, other
Challenges	No sub-codes were generated for this family. Instead, we applied this code family inductively on text where the authors expressed challenges with AI ethics education for **context**.	N/A
Quantitative assessment	These codes captured different quantitative mechanisms that authors used to **assess** students. These were adapted from (Hess & Fore, 2017).	Comparative pre/post, exam, instrument, rubric, survey
Qualitative assessment	These codes captured different qualitative mechanisms that authors used to **assess** students. These were adapted from (Hess & Fore, 2017).	Artifact assessment, interviews, observation, survey, written

Table 2
Frequencies of intervention papers by their strategy of intervention implementation, with the distribution of target population.

Intervention strategy		Duration of intervention	Target Population*	
Standalone course	5	4 full semester 1 full academic year	University	4 (/10) – 40%
			K12	1 (/11) – 9%
			Other (teacher, professional, public)	1 (/6) – 17%
Within course	4	1 one-class 1 three-week 1 one-semester 1 multiple courses in K12 year	University	3 (/10) – 30%
			K12	1 (/11) – 9%
			Other (teacher, professional, public)	0 (/6) – 0%
Extracurricular (workshop, summer camp, after-school)	9	1 one-day 4 one to three weeks 4 three months	University	1 (/10) – 10%
			K12	7 (/11) – 64%
			Other (teacher, professional, public)	2 (/6) – 33%
Other (Syllabi review, scoping survey, teacher interview)	7	3 one-course 4 full curriculum	University	2 (/10) – 20%
			K12	3 (/11) – 27%
			Other (teacher, professional, public)	3 (/6) – 50%

Note: Target population is not mutually exclusive (a paper can target more than one population), and percentages are calculated by [# within strategy]/[# total target population].

resources challenges.

In the proceeding sections, we will pay attention to how these authors tested AI ethics interventions while grappling with fundamental challenges. In as much, we will now review three emergent themes in AI ethics education.

4.3. Progressive pedagogies and advanced instructional methods

Are traditional teaching methods sufficient to prepare the next generation of AI users, developers, or policymakers for the societal challenges in the future of AI? If we consider the traditional "sage on a stage" model of traditional education, it is hard to imagine students learning the complex applied skillsets necessary to use and design AI systems. Moreover, the term and field of "AI" is conceptually complex. In our analysis, AI was framed inherently as a socio-technical-legal problem with apparent social and ethical dimensions (rather than purely mechanical or technical constructs). Although a common definition could not be found, we position this as *definitional flexibility*: enabling educators to adapt course material to their specific student contexts. In the following theme, we will understand what progressive pedagogies are being used that move beyond the lecture, notes, textbooks, and exam format of traditional instruction.

A rising challenge in AI education is how to design a course that stays relevant over time. "*[T]he rapid rate of progress in AI research and development [raises] the question of how we can ensure that at least some of the course content remains relevant to the students also in the future.*" (Tuovinen & Rohunen, 2021, p. 8). To do this, we saw authors choose active learning with contemporary examples and collaborative discussions rather than relying on lecture or text material alone to help prepare students for the future. For instance, Zhang et al. (2023, p. 302) recognized AI's multi-stakeholder nature inherent with bias, and effectively integrated content with pedagogical delivery for their students' age range:

Table 3
Key Findings for Each Code Family over 2018–2023 AI ethics education papers.

Code family	Key finding
Metadata	The majority of AI ethics education research is conducted by universities, with significant gender diversity in authorship and moderate international collaboration. The interventions varied in format and duration, with standalone courses and within-course deployments primarily at universities, while extracurricular activities targeted K12 students.
Target population	Despite creating space in the codebook, we found no interventions focused on AI experts, business leaders, policy makers, or the general workforce, and only one paper discussed public literacy beyond educational settings.
Ethics learning goals	Authors rarely stated explicit learning objectives, but we found that instruction focused on ethical reasoning. Deliberate practice of ethics communication and self-confidence was largely out of focus.
Educational context	Academic papers were not suited to reliably infer the components of formal and informal learning environments, but most interventions took place in a formal environment with casual instructional delivery.
Ethics topics	AI ethics educational interventions taught students about broad and societal ethics concepts through conventionally technical ethics concepts; Treating ethics more broad than narrow.
Instructional Delivery	The interventions were predominantly online and hybrid, but it is likely that the COVID-19 pandemic caused authors to transition in-person interventions to online and hybrid approaches, but some in-person interventions were still reported.
Pedagogical techniques	Progressive hands-on pedagogical methods were used to engage students in AI ethics activities, often using case-studies and group projects to facilitate instruction. Nearly all interventions leveraged a form of discussion-oriented learning.
Challenges	Authors expressed challenges with (1) AI complexity, (2) ethics obscurity, (3) educational principles, and (4) human resources while trying to implement their AI ethics interventions. See Table 4.
Quantitative assessment	Exams were rarely used to measure student learning, but comparative pre-/post-test instruments focused on testing the effect of the intervention with instruments to capture AI literacy and AI ethics constructs.
Qualitative assessment	Authors assessed students' artifacts to capture how students interact with AI ethics activities but often did not go into detail about specific student characterizations. Authors captured natural interactions of students engaging in AI systems.

They first create algorithms for making the best peanut butter jelly sandwiches and compare the algorithms. Then they bring their understanding of stakeholders like doctors, parents, and classmates (e.g., doctors probably care more about nutrients over the taste of the sandwich, classmates may care more about the taste, and parents may care both), to determine what "the best" peanut butter jelly sandwiches would include based on the perspectives of the stakeholders. Afterward, they discuss how the algorithms of "the best" peanut butter jelly sandwiches made by a parent would differ from the algorithms made by a doctor or a teenager.

While this example is targeted to young learners, the same principle of making AI material relevant to the students' lives can be encouraged to foster learning and engagement about complex social and ethical problems caused by current deployments of AI. After all, a day-to-day practice of ethics occurs within the lived experiences of students' lives.

Across our dataset, we witnessed higher levels than we anticipated for progressive pedagogies and advanced instructional methods such as hands-on learning activities, discussions, case study activities, and group projects. As the **most coded pedagogical technique**, authors engaged students in AI ethics material via hands-on activities. This included actively *using* AI-powered tools (e.g., Google's Teachable Machine in Williams et al., 2022, p. 12) or *creating* AI tools through code blocks, training AI models, or Raspberry Pis (e.g., Micro:Bit microcontroller and vision sensors on a robot in Eguchi et al., 2021, p. 157). In their K-12

Table 4
Core challenges in AI ethics education.

Challenge	As Represented in Papers
1 AI complexity	Tuovinen and Rohunen (2021) highlight the simple fact that AI changes quickly and that it is hard to pin down what to teach *now* that will remain relevant in the *future*. Additionally, the technical nature of AI systems asks educators to consider what *is* and *is not* relevant to include when teaching AI to non-technical audiences (Eguchi et al., 2021; Stoyanovich, 2022; Van Brummelen et al., 2021).
2 Ethics obscurity	Paired with the complexity of ethics, authors face the choice between teaching ethics *content* and ethics *skills* (i.e., do we teach ethics frameworks and principles or foster cognitive and behavioral tendencies for ethics). Moreover, when introducing students to AI ethics, educators must know how to show the translation from technical into sociotechnical as decisions in AI systems have broad social implications (Stoyanovich, 2022).
3 Educational principles	Eguchi et al. (2021), van Brummelen et al. (2021), and Hod et al. (2022), among others, express difficulty in correctly pacing their intervention, contextualizing it for different cultures, and identifying the appropriate pedagogical principles to deliver content.
4 HR challenges	It is hard to find the faculty and staff to teach. Sourcing material is challenging (Gong et al., 2020), finding the time to teach it might not be realistic (Garrett et al., 2020, pp. 272–278), and current teachers have limited experience (Zhou et al., 2022). However, as many authors express, teaching AI ethics is everyone's job, and educators for K12, university, public, or workforce settings are responsible for the development of responsible AI systems (Stoyanovich, 2022).

workshop, Van Brummelen et al. reported that "*The most frequently stated reason for engagement was the hands-on material–specifically the coding tutorials and group activities.*" (2021, p. 5).

Students do not even have to code, and one creative example of an aligned hands-on activity can be found in Zhang et al. (2023) where middle school students learn about logic systems by building a decision tree to sort macaroni from penne and classify twisted pasta versus flat pasta. That is an age-sensitive and effective way to connect content matter (AI logic systems) with pedagogy (hands-on activity) that connects students with the material. In university, students can think more abstractly and still engage in hands-on material; For instance, taking the Montreal Declaration for Responsible AI framework and analyzing AI applications under its constructs to bridge knowledge to practice (Taylor & Deb, 2021).

Generally, we found these educational researchers taking innovative liberties in their instructional design to keep up with evolving AI trends, whether by physical hands-on tasks, conceptual knowledge-building with emerging AI ethics frameworks, or synthesizing computer science material with engineering challenges and interdisciplinary awareness. With respect to RQ1, this spells out positive innovation in the future of AI ethics instruction. Approaching RQ2, it appears much of the instruction was informed by computing standards like hands-on coding and engineering practices like building and designing AI systems. Further on, we will examine how instructors use case studies and projects to elicit how AI is developed outside of the classroom.

4.3.1. Case studies, group projects, and discussion-oriented methods

Within the trend of progressive pedagogical techniques, we saw authors either use case-studies or group projects to engage students in educational material. With the use of case-studies, we saw authors argue for either (a) **creative and fictional** case studies or (b) **"authentic" and non-fiction** case studies. For instance, Forsyth et al. used short stories about AI with additional multimedia activities to explore AI ethics. "*The stories, which sparked an emotional connection to the different issues, were complemented by non-fiction texts and videos to ground the experience in current events.*" (2021, p. 2). They later go on to analyze how the stories heighten personal relevance to the students, and how an emotional

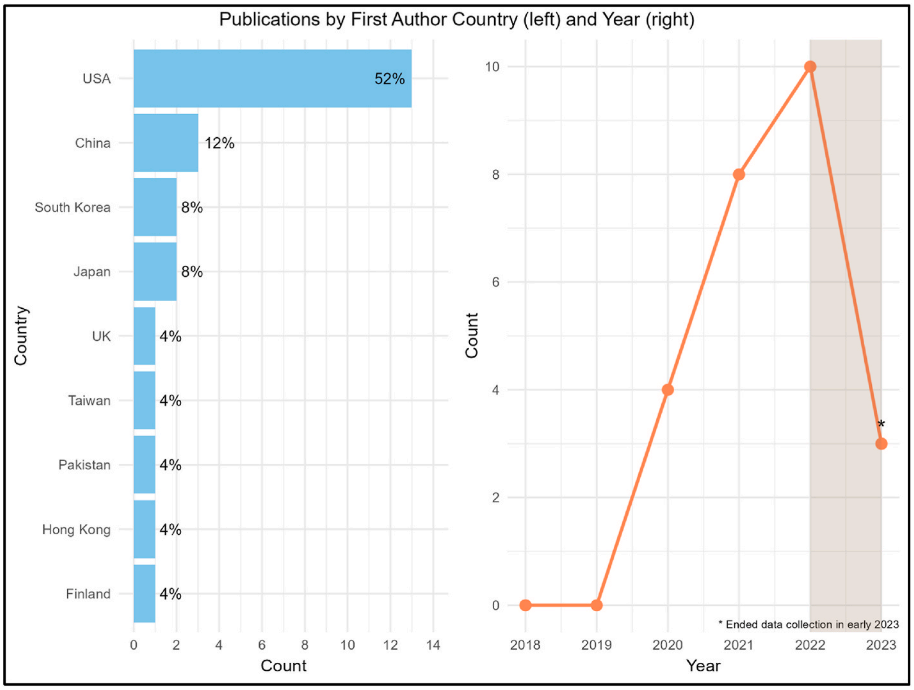

Fig. 2. Count of papers by first author's affiliation country and count by year published.

connection can facilitate learning transfer from the classroom to real-world scenarios. In Ng et al., students were coming up with creative stories themselves. The *digital story writing* (DSW) pedagogy pushes students to condense complex aspects of AI into digestible stories. As Ng et al. (2022, p. 3) write:

It is a powerful cognitive development tool to combine language, visual, and digital representation could enrich how students used images, sound, and movie editing techniques to express causality and development from one scene to another. During DSW experience, students have the learning opportunity to apply their prior knowledge, research, reflect from their daily life experience (Boase, 2008), explain and construct a finalised story.

Non-fiction, or, "authentic" case studies can complement creativity, as well. For instance, in an article focusing entirely on the effects of case-study methods to teach AI ethics, Hishiyama and Shao wrote: "*By dealing with real problems, we have developed teaching materials for discussion with the help of complex and realistic cases [such as deepfake news]*." (2022, p. 228). Similarly, in the re-representation of K12 teachers, Kim et al. (2022, p. 6080) write that:

Teachers highlighted the use of authentic tasks that allow students to construct and apply standard-driven knowledge to solve real-world problem need to be foregrounded. Teachers developed various learning activities that make connections between subject-area knowledge and real-life problems through the student-AI team's task-focused interactions.

Whether fiction or nonfiction, the act of solving case-study problems exercises creativity. Students must transpose the case study material into a real-world scenario and work within its constraints to come to a solution; This may include synthesizing course material, developing multimedia artifacts to explore AI ethics issues, or engaging students in teamwork to practice group project problem-solving. An outstanding question, though, is when to use creative case studies or authentic case studies and which cognitive skills each method promotes.

The use of group projects appeared apart from case studies. Interestingly, case studies and group projects were not found in use together, and we suggest that this split is because of the intuitive understanding of a *project* as something that is created fresh, rather than recycling and studying existing stories, materials, and news (i.e., case studies). Ideally, project-based learning occurs in context-specific active learning environments with social interaction (Kokotsaki et al., 2016), and in our analysis, we see these group projects as attempts to encourage collaborative learning.

For example, in a repeated *t*-test analysis, Shih et al. (2021, p. 11) found that "*combining hands-on activities and group work can help non-engineering students enhance their perceptions of AI issues and strengthen their awareness of interdisciplinary collaborative learning.*" Collaborative group projects synthesized course material and taught students how to collaborate with others to build AI/ML products for societal problems (Chklovski et al., 2020, pp. 34–35; Van Brummelen et al., 2021). Julia Stoyanovich pinned her Responsible Data Science course on project-based learning. At the apex of teaching responsible AI via comics and self-reflective memos, her "*course project pursues the broad learning goal of making [automated decision systems] interpretable […]. Adhering to constructivist principles [where] students work in teams of 2 to audit an automated decision system (ADS) of their choice.*" (Stoyanovich, 2022, p. 7). In conclusion, we suggest that the nature of AI–as a multidisciplinary and *multisystem* endeavor–warrants these project-based methods more than most fields. To integrate case studies and group projects, we propose, for example, to take case studies of an AI system "gone wrong" and build projects to find more secure solutions and elicit student engagement.

Even without a substantive *project,* discussion-oriented methods were used to facilitate collaborative learning. As the **second-most found pedagogical method** in these papers, discussions were strung through both case studies and group projects and were not found to be dependent on a more central pedagogy. Naturally spread through interventions, students would, for instance, "*discuss the ownership of machine-generated art*" (Zhang et al., 2023, p. 300), or, after case study analyses engage in a class period of discussion (Hishiyama & Shao, 2022, p. 4). If there were theoretical material, a "*supplementary session would then be to have the students discuss their thoughts with peers …*" (Tuovinen & Rohunen, 2021, p. 10). Using discussions in classes is nothing new, of course, nor do we have perfect insight into *how much* of an intervention leverages the social, peer, and collaborative aspects of discussions. But, since nearly all papers that detailed an intervention include an aspect of discussions, we are likely to see this trend continue regardless of other instructional methods or content matter. Moreover, we suggest this helps promote practices of multi-stakeholder participation in AI development where students build habits of discussion that can contribute to responsible public discourse in the future.

Using more progressive, or advanced, pedagogical methods may be expected from our corpus of AI ethics education articles. After all, the academic publishing industry–and the passions of authors writing about education–are likely not to be reports of lectures, textbooks, and exams. However, to summarize, authors were less traditional than we anticipated. For another example, outside of projects, case studies, or discussions, progressive pedagogy could be where students *"collectively play a game to understand societal consequences of AI-generated media."* (Zhang et al., 2023, p. 11). Multiple authors engaged students in creative writing, drawing, musical composition, or comic-strip storyboarding to elicit AI ethics skills and content (Kim et al., 2022; Ng et al., 2022; Stoyanovich, 2022). While not exhaustive, the point is that these novel uses of AI ethics content in the classroom can engage students in multiple forms of learning. They use rote knowledge, imagine ethical futures (Young & Annisette, 2009), and synthesize problems and solutions into different forms of media. This may push students to more deeply consider the forms of AI ethics outside of the classroom. Meaning, AI ethics exists in the real world, public and professional life, and the fundamental problems will come in multiple forms and likely need to be solved in multiple forms.

4.4. More macro-ethics than narrow-ethics

Technical fields like engineering and computer science are often grounded in a positivist paradigm where research and innovation are driven by observable, measurable activity. This approach is valuable for evaluating the efficacy of scientific advancement or innovation. However, it has also faced criticism for promoting a mechanistic view of humans, society, and ethics (an argument akin to Horkheimer and Adorno's *Dialectic of Enlightenment*). Consequently, attempts to define and control ethical impacts can become formulaic and algorithmic, reducing society to metrics. In AI, these can help hold algorithms accountable (Kroll et al., 2017), but can also create a narrow conception of ethics that fails to consider broad societal and environmental risks (Raji et al., 2021, pp. 515–525). For these challenges, we found that authors positively considered both narrow and broad conceptions of ethics. As an exemplar: "*this [learning] module takes a lifecycle view of responsible data science, as a step towards a more holistic (rather than reductionist) treatment of technology ethics.*" (Stoyanovich, 2022, p. 6).

4.4.1. Risk, bias, and privacy

Out of all ethics topics coded (Table 1), the risk of AI being used for potential harm or danger was the **most coded ethics topic** in our dataset, we saw that authors taught ethics by discussing how the nature of AI causes societal concern. Sometimes, these were standalone, general, and unspecific focuses like "*students [should recognize] that AI has potentially negative consequences* …" (Ottenbreit-Leftwich et al., 2022, p. 10). Other times, the risk of AI was coded alongside topics like the risk of privacy, consequences of autonomous robotics, dangers of algorithmic bias, or proliferation of misinformation. In Zhou et al.'s (2022, pp. 337–343) review of international trends in AI education, they found that AI ethics interventions often covered the negative impacts of AI in its privacy, security, or data bias risks to explain how technical constructs can impact society. We found that authors considered the *risk of AI* as a broad catch-all to talk about ethics. They bucketed multiple societal and human concerns as a risk to the development of AI, sometimes without becoming specific to certain ethical harms.

The **second most coded ethics topic** in our review was the tendency of a system to favor one group of people over another: *bias*. We saw that authors shuffled between the perspective of technology and the perspective of societal harm. Meaning, technical data bias and the impact on human discrimination were considered as one, not separate. For instance, bias may be introduced to students through technical concepts of data bias and algorithmic bias but then explained in terms of large-scale societal inequalities. Ng et al. introduced this concept "*through the hands-on simulation of AI applications, students could recognize the importance of data bias and it is important to avoid unjust impacts on people* …" (2022, p. 7). A similar representation was found in Van Brummelen et al. (2021), and generally represented the approach we saw in these articles: technical topics being introduced to students but not in isolation of the ethical or societal impacts.

As the most technical code that shared a significant portion of our dataset's focus, privacy was the **third most coded ethics topic**. Privacy is a critical concern for AI–and one of the most focused aspects of digital (Dempsey et al., 2022). Privacy over online material and data protection for intellectual property can be seen as a social topic just as it is a technical construct in AI development. Nevertheless, in our dataset, privacy was often referred to as data protection in a more technical fashion than social or human-rights conceptions. For instance, in Stoyanovich (2022), students worked with synthetic data generators to demonstrate how to protect privacy. And Forsyth et al. (2021, pp. 1–2) included class activities on data privacy with social media, cell phones, tracking devices, to demonstrate differences between safety and privacy. Additionally, in both the given examples, these authors engaged students with privacy with hands-on and active activities like Google's Teachable Machine, MIT's Moral Machine, and the Jugi chatbot (Forsyth et al., 2021, pp. 1–2; Stoyanovich, 2022). Here, privacy is considered a technical control that can be moderated through systems development. However, the risk of violating privacy (the *risk of AI)* can bridge the technical to the socio-legal-technical.

4.4.2. K12 and higher education's distribution of ethics topics

Educational needs for young children and young adults will naturally be different. It's assumed that young learners (i.e., K12) need to start with more digestible topics to set foundational knowledge whereas older learners (i.e., higher education) can tackle complexity, nuance, and the multiplicity of ethical situations. To understand the differences in ethics topics in K-12 versus higher education, we created Fig. 3 to visualize the wider spread in higher education versus the tighter focus in K-12. In this radar chart, each edge node is an ethics topic, axes represent frequencies, and the purple and orange polygons visually represent the different frequencies between K-12 and higher education.

We suggest two explanations for this difference: heterogeneity of higher education contexts and/or the level of sophistication of K12 and higher education learners. First, higher education spans multiple majors, different university agendas, and more faculty discretion in choosing topics. Contrary to K12 which may be more confined to standard curricula and less free to experiment with complex ethics topics. Second, higher education learners may simply be able to conceptualize a wider array of topics related to AI, whereas K12 learners may need to start with easily-digestible foundational topics. This distinction may help elucidate an opportunity for K12 to diversify AI ethics topics for young learns, or, to suggest that AI ethics research should focus on how AI ethics should be contextualized for young learners. As shown in this data, there may be a gap in making ethical theories and ethical codes relevant or making transparency and explainability, digestible to the K12 student.

With consideration of RQ1, we suggest that AI ethics scholars, broadly, might in fact be constituted of two different populations: higher education researchers and K12 researchers. These subsets of scholarship both broadly treat ethics as holistic rather than reductionist, but they might have different conceptions of which AI ethics topics are relevant. For example, take the topics of inequality, power, fairness, and diversity, equity, and inclusion (DEI)[4]: Both ethics topics were coded the same number of times, but the former was considerably *more* present in higher education than K12 whereas the latter (DEI) was considerably *less* present in higher education than K12. As we analyzed this, we saw that inequality, power, or fairness included more discussions about

[4] See Appendix A for the operational differences between these two codes.

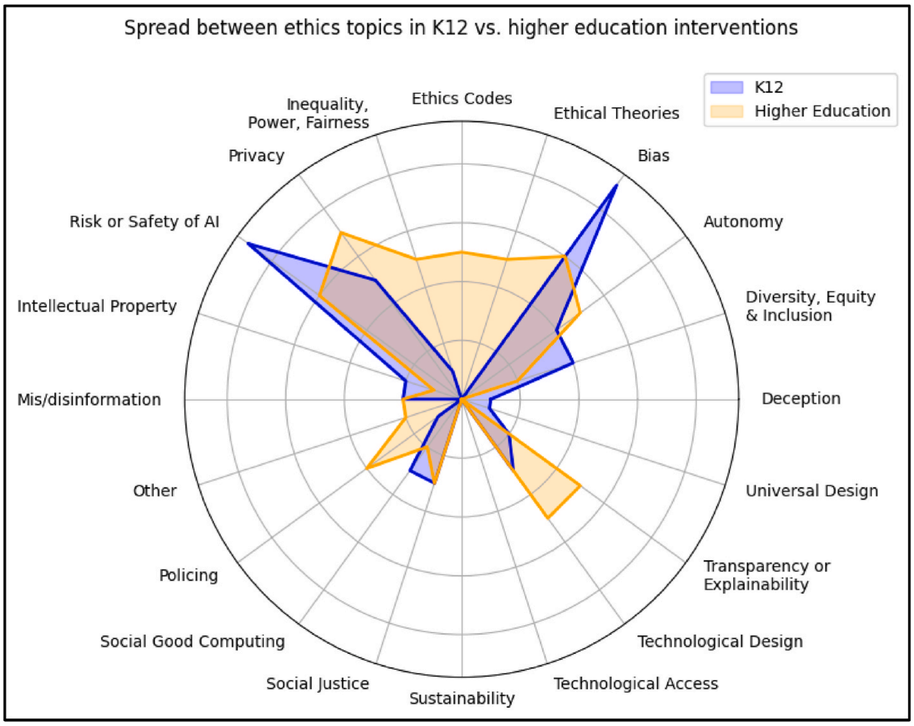

Fig. 3. A visual of the spread between ethics topics in K12 interventions versus higher education interventions. Labels from Barkhuff et al., 2024.

algorithmic bias, discrimination, and fairness with respect to how machines learn from data. Alternatively, we saw that DEI codes included more discussions on race and gender discrimination and highlighted AI's general impact on diverse people and communities. This is not a comprehensive analysis, but with Fig. 3, we suggest it may be the case that scholarly communities are conceptualizing AI ethics differently in K12 and higher education settings.

4.5. Research evaluation preference over student assessment of learning goals

We found authors face a challenge to balance practically implementable, but robust, student assessment methods to support students all the while incorporating measures for research evaluation necessary for academic publishing. The distinction is nuanced, but in general, we saw research evaluation valued over student assessment. "*The danger here is that we end up valuing what is measured, rather than that we engage in measurement of what we value.*" (Biesta, 2009, p. 43). In other words, we should be cautious to not let accessible research evaluation be interpreted as a valuable assessment of student learning. For instance, as we will show, authors *implicitly* encouraged students to put their knowledge to use with critical analyses of AI systems, but this was not often *explicit* and thus not easily articulated and measured. In the proceeding theme, we first detail where we noticed lackluster assessment, then what assessment methods were used, and finish with how learning goals were incorporated (or not) into educational design.

4.5.1. Trivial necessity of assessment

When relying on summative assessments in an intervention, it will be expected that students learn *something* about the course material. However, this may be due to a natural maturation effect of testing. We found that summative assessment was often used rather than formative assessment, despite the authors engaging in progressive modes of learning, and a holistic conception of ethics (Themes 1 and 2). For instance, students in Ng et al.'s (2022) intervention wrote creative stories to learn about a wide array of ethics topics but were assessed by interviews and a pre-/post-test instrument measuring AI literacy. The authors conducted a baseline pre-test on AI literacy, spent 3 months on AI and AI ethics, and conducted a post-test and found positive change in AI literacy. We are left feeling that a positive change *should* be expected–a maturation effect over that long of a time. As such, stating positive outcomes feels trivial, but a necessity of completing these interventions.

Similarly, in Green's evaluation of their AI ethics course, they write that "*In [summary], the projects demonstrated that the students were able to analyze some ethical issues relevant to their [AI] agent and then apply ethical principles (explicitly or implicitly) to the implementation.*" (2021, p. 522). However, the AI ethics skill of applying principles to practices is not just a binary, yes/no, application; It is a complex practice, and assessment should provide feedback on how well this is done, rather than merely if it is done. Moreover, Green structured their course based on stated learning objectives but did not revisit them to structure the assessment of student learning. These exemplars are not listed to critique them, but rather to highlight a necessary pitfall of assessment. We understand that this may, of course, be because there is only so much space in a publication, and assessment was chosen as the short end of the C.A.P. alignment stick. Still, these two examples suggest that the trajectory of AI interventions will remain unaligned–innovating over assessing–without a substantive focus on student-centered learning.

4.5.2. Assessment methods in AI ethics interventions

As mentioned previously, it was often hard to differentiate assessments for students (given to students as feedback) or assessments reported for publication. Nevertheless, assessment (of any kind) was split between qualitative and quantitative methods. While 20% of papers did not detail any type of assessment, 40% used both qualitative and quantitative techniques, 24% used only qualitative techniques, and 16% used only quantitative techniques. The most common qualitative technique was *artifact assessment* followed by *written assessment* and *classroom observation*. The most common quantitative technique was a *comparative assessment* (pre-/post-test) followed by *surveys* and *rubrics*.

Different assessment techniques seemed to pair with different pedagogical approaches: Papers with hands-on pedagogies predominantly used mixed assessment mechanisms, papers with group projects never used quantitative mechanisms, and papers using case-studies were assessed with quantitative mechanisms more often than not. Moreover, papers that discussed the ethics topics of the *technological design* and *transparency* of AI systems never used quantitative assessment, while papers that discussed the *privacy* of AI predominantly used quantitative assessment.

Due to the complexity of AI ethics and the multi-modal methods that were used to teach AI ethics, we found it was uncommon for only a single type of assessment to be used. For instance, Forsyth and colleagues used "… *pre-/post-surveys, observations, and analysis of student artifacts to evaluate the [educational] program's effectiveness and in particular, the use of short stories in heightening students' awareness and understanding of AI.*" (2021, p. 2). *Artifact assessment*, as the most common qualitative assessment, was coded broadly and encompassed the practice of evaluating something created by the students. Rubrics were used in conjunction to evaluate the projects, or artifacts, based on pre-existing criteria and objectives. The use of artifact assessment, as found, can capture learning in a more naturalistic manner. Just as the pedagogical methods used hands-on activities, group projects, and case studies to emulate the real world, the authors demonstrated the recognition that assessment needs to occur in the same manner. Only two articles mentioned the use of an exam (Chklovski et al., 2020, pp. 34–35; Taylor & Deb, 2021).

On the quantitative side, *comparative assessments* were coded when authors wanted to assess certain metrics for AI ethics or AI literacy. For instance, Shih et al. (2021) tested for increased awareness of AI ethics issues, Williams et al. (2022) compared student understanding of AI concepts such as embedded ethics, and Bae et al. (2022) assessed AI ethical competence in students' computational thinking. In sum, the authors deployed multiple forms of assessment. However, most papers did not go into much depth about their chosen assessment and how, or if, it provided formative feedback to students. Nevertheless, we are left with an opportunity for future AI ethics interventions to pay more attention to these gaps and treat assessment as something more than mere research evaluation.

4.5.3. Ethical learning goals and objectives

Educational interventions designed and deployed with specific learning goals in mind can promote instructional alignment that can improve educational effectiveness (Cohen, 1987; F. Martin, 2011). However, only three authors *explicitly* specified learning objectives. First, Eguchi et al. (2021) organized their material around AI4K12's "Five Big Ideas in AI" (Touretzky et al., 2022). They stated goals ranging from understanding the basic mechanics of AI to understanding how AI is a multi-stakeholder interdisciplinary problem that has a large impact on the world. Second, Williams et al. (2022) structured their curriculum around three learning outcomes that scoped from understanding technical knowledge, critical thinking about society, and applying AI knowledge. Last, we found that Hod et al. (2022) defined three learning objectives for the *character of the student* such as promoting multidisciplinary dialogue, responsible AI literacy, and professional responsibility. Overall, these objectives range from the technical aspects of AI, its socio-technical components, and ethical scenarios to consider.

Since only three articles defined explicit learning objectives, we coded the remaining articles by inference to capture abstract and higher-level ethics education goals. These goals were (a) understanding intuition's effect on ethical decision-making, (b) learning knowledge or theory about ethics, (c) promoting ethical reasoning skills, or (d) improving communication skills and confidence within ethical situations. See Appendix A for operational definitions.

We found *ethical reasoning* as the most intended ethical learning outcome and coded it 15 times in our dataset. Sometimes authors made explicit reference to ethical reasoning, like exploring "*[the] role of social norms and sacred values in* ***ethical reasoning***," (N. Green, 2021, p. 3). Others alluded to ethical reasoning indirectly. For instance, Williams et al. described how "*students can think critically about the potential benefits or harms of AI systems and their impact on stakeholders*" (2022, p. 335). Both of these quotes approach an encompassing view toward ethical reasoning as a focus on applied skills that help students make ethical decisions.

Then, *knowledge* was the second-most coded ethics learning goal. Often coded alongside other goals, like ethical reasoning, Hod et al. described it where students' objectives were to "*possess introductory knowledge and skills to oversight, audit, and steer AI systems through their life cycle.*" (2022, p. 37) alongside performing stakeholder analyses and identifying consequences and harms. In this, the goal of the intervention is to be aware of ethical issues and conduct procedures to reason through ethical problems in practice, blending both knowledge and reasoning.

The goals of promoting *communication* and *fostering intuition* about ethics were coded nearly at the same frequency. First, in our dataset, building communication skills around ethical topics fell into two subcategories: (1) written or oral communication for the sole intention of practicing communication, like a presentation, and (2) communication with peers, like a discussion. Second, fostering intuition about ethics was coded students engaged in *stories* or *scenarios* that may relate to personal topics from their lived experiences. For example, Ng et al.'s (2022) story-writing pedagogy has students connect their personal experiences with AI ethics through writing stories, and engage their personal experiences and emotions in future scenarios. Overall, we found *communication* as the externalization of the AI ethics practices process and *fostering intuition* as the internalization of AI ethics practices, both being necessary bookends to mere ethics knowledge and reasoning.

In conclusion regarding the above three themes, we have sketched a picture that (1) there are interesting and creative ways that articles synthesize advanced methods to teach advanced content, (2) the type of content they are teaching is somewhat broad, not narrow, which is probably good in the attempt to move away from strictly technical ethics, but (3) more work should be done to provide formative assessment to students about the effectiveness of techniques and indicators of change. While many authors captured societal north-star objectives with their AI ethics, more work can still be done about what is truly important about AI ethics and how we can move the needle toward naturally occurring ethical behavior.

5. Discussion

This paper analyzed AI ethics education from the dimensions of content, assessment, and pedagogy of real educational interventions. Our results sketch a picture of the early years of substantive AI ethics education to infer the trajectory of strengths and weaknesses of the field. We highlight the current trends in using creative pedagogical methods for teaching AI ethics, with greater focus on broader concepts of ethics over narrow ethical principles, and the lack of formative assessment tools to evaluate the cognitive side of educational effectiveness and alignment with the learning objectives. These findings pose implications for both researchers and practitioners in the field of AI ethics education.

5.1. Implications for researchers

Our SLR looks directly at the practices of educational researchers, and as such, is an analysis *of* educational researchers. In this, we draw multiple implications for researchers in AI ethics education and complementary fields. For instance, situating our findings in the learning sciences can help us understand the cognitive demands of *learning* AI ethics and *being* an AI ethicist. In turn, informing better curriculum design and more effective pedagogical techniques for AI ethics material.

5.1.1. Integrating non-traditional learning methods in classrooms

Our analysis found that authors incorporated new material in the classroom largely through co-constructive learning activities (e.g., case studies, group projects, and discussions). Some pedagogies, like creative story writing (Ng et al., 2022), game playing (Zhang et al., 2023), or comic-strip creation (Kim et al., 2022) may be uniquely suited to teach computer science and AI. For instance, students can engage with controversial societal topics in a safe and "fictional" manner while engaging the emotional cognition of the student (Burton et al., 2018). One of the approaches we did not see, though, was the use of role-playing. This pedagogy has recently gained attention, and in the landscape of AI that involves competing stakeholder groups (Sadek et al., 2023), role-playing can provide students with a valuable practice where students act as stakeholders and talk directly to competing interests and complex emotions (Avin et al., 2020; Shapiro et al., 2021).

To further investigate these innovative methods in educational settings, we recommend following the design-based research (DBR) methodology. DBR improves educational practices by providing a framework to test and measure classroom strategies while maintaining a naturalistic environment (Barab, 2014; Magana, 2022).[5] Under a DBR methodology, researchers can continue the use of innovative pedagogy, while still being rigorous and systematic in its evaluation. This affords AI ethics education to account for the new developments in technology while maintaining educationally sound assessment practices for student learning.

5.1.2. Situating AI ethics in computer and engineering education

AI shares many of the same traits of computing and engineering curriculum but we found a few key differences worth mentioning. First, the ways of integrating ethics content in AI curricula appear to differ, slightly, from computing curricula. If we look at Brown et al.'s (2024) review of computing ethics, the conception of ethics was more rooted in philosophical theories, codes of ethics, and mechanical views of principles. In our findings, however, we rarely saw philosophy discussed, and codes of ethics were used as a means to an end for ethical reasoning. Second, it appears that computing ethics relies more on lectures and readings, where discussions and writing assignments are secondary (Brown et al., 2024). Our review, on the other hand, did not emphasize lectures or reading material quite as much. Nevertheless, we found one key point of overlap: authors seemed to value research evaluation over formative student assessment. Most computing ethics education programs were evaluated with "simple" measures of student satisfaction—easily reportable—rather than students' cognitive changes in learning (Brown et al., 2024).

We also reviewed Hess and Fore's (2017) SLR on engineering ethics interventions to find differences and similarities. The primary difference was how engineering contexts focused more on rote teaching of codes and theory whereas AI used more co-constructive pedagogy. However, both engineering and AI interventions had similar objectives that focused largely on ethical issue-spotting along with ethical reasoning skills. Then, we saw engineering contexts also rely primarily on student self-report metrics to evaluate the engineering ethics interventions (Hess & Fore, 2017).

Regardless of pedagogical or curricular differences, AI is naturally a part of engineering and computing. From the planning and development of AI to the deployment and monitoring of AI, different stakeholders from engineering and computing disciplines play an important role in the responsible development of AI (Lu et al., 2024). As such, paying attention to research in these respective disciplines will still promote the responsible development of AI (Abulibdeh et al., 2024; Southworth et al., 2023). We find that there is an opportunity to explore how existing engineering and computing education frameworks can be adapted to uniquely account for AI ethics principles and challenges—rather than reinventing the wheel for AI.

5.1.3. Aligning assessments with learning goals

An elusive assumption underlying the three themes is that classroom learning translates into behavioral change and real AI ethics practice. While we found that current pedagogical practices aim to do more than just transmit knowledge (by incorporating creative and reasoning-focused methods), we found it unclear whether these methods effectively lead to ethical behavioral change. Just because there is an ethics intervention does not mean students will accept the ethics intervention. For instance, when grades (rather than ethics) become valued, when technical material distracts from ethical material, or when students' preconceived notions of the world do not match the material, students may resist the ethics intervention—even if they get "good grades" in class (Padiyath, 2024). The literature on changing behavior through interventions is controversial and interventions that focus on knowledge and general skills have a negligible effect (Albarracín et al., 2024), though it is hard to find feasible strategies that do have an effect.

Part of the problem here is defining the goalposts; what *is* ethical change? Relatively easy things to measure, such as changes in knowledge or reasoning capacity, may create an unintended consequence. To demonstrate, Zhong (2011) conducted three experiments and found individuals who were primed to deliberate more engaged in more unethical behavior. Indicating that training in ethical reasoning may increase unethical behavior.[6] This debate also extends to business, management, and economics fields that have scrutinized ethics education as an effective way to promote pro-ethical change (L. Wang et al., 2011). Unfortunately, though, moving away from reasoning and rationality has been met with controversy and disagreement (Clancy & Zhu, 2023). Because ethical decision-making in AI is multifaceted and context-dependent, it is difficult to develop measurement tools that can accurately gauge a student's ability to navigate real-world ethical dilemmas (Hagendorff, 2022a). What constitutes ethical behavior, and what contributes to behavioral change, is a topic under much uncertainty in AI ethics (*citation blinded for peer review*).

5.2. Implications for practitioners

5.2.1. Integrating an AI lifecycle approach in the AI curriculum

Our analysis of AI ethics educational practices has implications for the practitioners in the AI field as well. Education qualifies students to be workforce-ready, and in turn, shapes practices in the workforce itself; What comes out of education goes into industry. So, when we analyzed the 'risk and safety' of AI, we found that most papers focused on *deployment* and *usage* risks rather than the *planning* and *systems design* risks of AI systems. Of course, companies need to take AI ethics risks seriously at the start—during planning—rather than playing catch-up down the road (Chen & Ahn, 2022). Instructionally useful is L. Wang et al.'s (2021) overview of how AI impacts different stakeholders, at different times, and in different ways. This framework can model how the AI lifecycle—and its risks—are seen across sectors. In turn, hiring managers can recognize which students have AI literacy skills across the lifecycle, or only in the deployment and use of AI technology.

5.2.2. Operationalizing ethics in AI audits

Assessment in education is like auditing in industry. Conducting audits to assess the complex practices of AI against policies, regulations, and standards is an important role in the development lifecycle of responsible AI. Audits can prevent costly errors, reduce the risk of regulatory penalties, and enhance the overall trust and acceptance of the organization (developing and using AI) among stakeholders (Raji et al., 2020, pp. 33–44). However, current AI auditing frameworks are influenced by traditional financial auditing systems and mostly focus on *risk assessments* of AI systems (Schiff, Kelley, & Camacho Ibáñez, 2024) which may fail to capture the social costs of AI systems (Hagendorff, 2022b). As finding effective assessment metrics is hard in education, finding suitable AI ethics success metrics is hard in the industry. In both contexts, we face the pertinent challenges that feasible assessment often

[5] For a collection of research on innovate pedagogy see (R. K. Sawyer, 2022) and the older 2nd edition (K. Sawyer, 2014).

[6] See Haidt (2001) and Reynolds (2006) for foundational pieces revealing the secondary role of conscious reasoning in the ethical decision-making process.

Table 5
Summary of potential future research directions with suggested questions.

Domain	Example Questions	Context
Content on AI ethics	● What are unique considerations for AI education for children learners, young adults, or adult learners in AI ethics? ● What are pragmatic case studies that can help students learn the technical engineering challenges of ethics along with social challenges of ethics? ● How do industry practices in AI influence the AI ethics content matter in focus in education?	Preliminary understanding of AI ethics for children conducted by (G. Wang et al., 2024).
Assessment and evaluation of AI ethics education	● How can computational methods identify patterns of behavioral change in naturalistic settings outside the classroom to evaluate and metricize reasoning and communication skills? ● Which industry AI ethics metrics can correspond to practically implementable metrics to measure and assess in education? ● How are Higher Education policies incorporating assessment and measurement while spreading AI ethics across the curriculum?	Our findings suggested needing better ways to assess learning. Industry AI metrics are beginning to be catalogued (Xia et al., 2024) and translational work is needed.
Pedagogy and instructional methods	● Why is there a split between using group projects and case studies? ● When do creative instructional case studies prove more effective than more business-oriented case analysis? ● How are real workflows of AI professionals incorporated into classroom pedagogies?	Studies like (Griffin et al., 2024; Pant et al., 2024) qualitatively investigate AI professionals.

narrow down the effectiveness of such assessment. Policymakers, auditors, and industry professionals designing AI metrics should pay attention to this challenge and determine how to ensure the sociotechnical accountability of AI systems.

5.3. Limitations

While the findings and implications from the current study are informed from a systematic, detailed, and rigorous review of the field's literature, it is not without its limitations. First, in a systematic literature review, it is common practice to define the conceptual domains you are interested in (AI, ethics, and education) and create a set of keywords to search for these domains. However, we only initially subset for *AI Ethics* literature, and later filtered for education. This choice was to account for the ambiguous boundaries of *Education* and more reliably capture public training efforts, workforce development, and professional training initiatives. This approach let us cast a wider net and let us export an entire corpus on papers with any phrasing of AI ethics or responsible AI. Then, with manual exclusion, we could ratchet down a more directed educational scope.

Additionally, deductive qualitative content analysis is susceptible to "leaving things out" (Graneheim et al., 2017). We created our codebook from pre-existing taxonomies, theories, and concepts in the AI ethics and education spaces, and thus may have left things out that we did not look for. However, in response to this concern, we encountered several occasions where we were unsure whether to include or exclude data based on our codebook scope. During team meetings, we reviewed these cases and erred on the side of over-inclusion, rather than exclusion. Moreover, during thematic analysis, we let patterns and meanings arise naturally (as best we could) without forcing preconceived notions onto the data. This was also kept in check by our use of memoing and reflections. These tools helped keep the research team honest and treat the data as raw, pure, and emergent as possible. Naturally, we should recognize the qualitative nature of this study and the methodological limitations that go along with it but nevertheless, position the comprehensiveness of this qualitative work as a strength in the field.

Additionally, in our data, over 50% of the reviewed studies are based out of the United States, which implicates the general representation of research in AI ethics education. Another natural trait of our study's time period is the COVID-19 pandemic. As the emerging years of AI ethics education started to solidify in 2018–2023, the pandemic impacted educational spaces and settings. This may have impacted feasible learning strategies and assessment techniques available to the authors. Nevertheless, we did not explicitly observe any distinctive patterns associated with COVID-19, except for the transition from physical classrooms to virtual settings.

6. Conclusion

The demands for AI literacy, ethical use of AI, and responsible development of AI are growing day by day. Educational efforts for AI literacy and AI ethics will continue to become more concerning for educational institutions and workforce development programs. As such, we responded to a need to understand the trajectory of these efforts by assessing the educational effectiveness of AI ethics interventions through an analysis of its content, assessment, and pedagogical components. Our systematic literature review collected AI interventions during the early years of this field and carried out an in-depth qualitative analysis to explore the salient intricacies that have sizable implications down the road.

Future research can continue the promising and innovative research that took place during the early years of AI ethics education. While we found gaps in the educational alignment, many of the research programs inspire closely related work to be completed. We have compiled these fruitful research directions into Table 5 which provides example research questions to be answered in the future. As we continue to explore these directions in AI ethics education, it becomes increasingly clear that frameworks based on multi-disciplinary dialogue and multi-stakeholder engagement approaches are crucial to operationalizing ethics beyond the classroom.

In conclusion, we explored the trajectory of AI ethics education by observing how advanced pedagogical methods and non-traditional forms of learning are paired with broad societal views of ethics for students. With this, though, we saw evidence that led us to a concerning–or, skeptical–look at educational assessment and educational value. To maintain education's role in preparing responsible AI practitioners, we suggest future work across all AI, computing, and education should take a more concerted view of robust educational assessment and not conflate it with research assessment. We presented our research questions alongside key challenges faced by the field and key findings on each dimension of our conceptual codebook. In total, this led us to discuss the implications for both researchers and practitioners. We distill recommendations for both into *ways forward*, and suggest critically appraising how to translate industry demands of metrics to track responsible AI into educational material; Balancing best practices from the corpus of Education with the nascent demands of AI ethics will remain a challenge. All in all, we find the scholarly discourse on AI ethics education to spell a positive view of the field, but not without paying close attention to how misalignment can exacerbate AI ethics problems down the line.

CRediT authorship contribution statement

Lucas J. Wiese: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Indira Patil:** Writing – review & editing, Writing – original draft, Visualization, Validation, Investigation, Formal analysis, Conceptualization. **Daniel S. Schiff:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Alejandra J. Magana:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

Lucas Wiese reports financial support was provided by the National Science Foundation. Other authors do not declare competing financial interests beyond what is reported or personal relationships that could have appeared to influence the work reported in this paper. *Additionally,* internal institutional funding was provided by the Office of Research at Purdue UniversityPurdue. The authors declare that the manuscript is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. There are no additional declarations.

Acknowledgments

This work was supported in part by Purdue University's Office of the Vice President for Research and Partnerships and Purdue Polytechnic Institute through the Seed Funding for Academic Books and Monographs program. This work was additionally supported by the U.S. National Science Foundation under award # 2134667. The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of Purdue University, NSF, or the U.S. Government. Additionally, the authors wish to thank Adam Hafez for his support in initial data collection and Dr. Wei Zakharov for assistance with methodological study design.

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.caeai.2025.100405.

Appendix A. Complete Codebook

Table A.1 in this appendix shows the full technical codebook used in the analytical coding procedure during methods of this paper. The Code Family is the highest unit of organization, Code is the analytical unit that applies to the text, and Code Description assists the researchers with operational guidance to apply these codes in practice.

Table A.1
Full codebook organized by Code Family, Code, and Code Description/Guidance.

Code Family	Code	Code Description or Guidance
Metadata	international_collab	Whether the primary affiliations of two or more authors are located in different countries. Must have 2 or more countries.
Metadata	methodology mixed methods	If the paper itself, or research design as a whole, is mixed methods (both quant and qual).
Metadata	methodology qualitative	If the paper itself deploys qualitative methodologies in its analysis
Metadata	methodology qualitative	If the paper itself deploys quantitative methodologies in its analysis.
Metadata	research_design empirical	Whether the publication primarily focuses on empirical approach. Mark empirical if it is empirical but not intervention.
Metadata	research_design intervention	Whether the publication proposes an (educational, training, etc.) intervention. Most likely will be empirical, but if it is an intervention and empirical, mark it as intervention.
Metadata	research_design review	If it is a review paper like an SLR, syllabus analysis, meta-analysis, etc.
Metadata	research_design theoretical	Whether the publication primarily focuses on theoretical aspects.
Metadata	sector_civil-society	Refers to the sector of the organization the first author of the publication is affiliated with. University, civil society, industry, government, unclear.
Metadata	sector_government	Refers to the sector of the organization the first author of the publication is affiliated with. University, civil society, industry, government, unclear.
Metadata	sector_industry	Refers to the sector of the organization the first author of the publication is affiliated with. University, civil society, industry, government, unclear.
Metadata	sector_university	Refers to the sector of the organization the first author of the publication is affiliated with. University, civil society, industry, government, unclear.
Metadata	sector_unclear	Refers to the sector of the organization the first author of the publication is affiliated with. University, civil society, industry, government, unclear.
Metadata	Woman_NB	Whether any of the authors of the publication are women or non-binary.
Target_Population	AI_experts	(industry) AI Creators: This group comprises individuals undergoing specialized training to become AI experts, such as those pursuing specialization in AI or working as AI Scientists. They are the creators and innovators in the field of AI.
Target_Population	business_leaders	Business Leaders: This refers to CEOs, executives, managers, or other individuals who have a localized impact on a specific business or firm. They may not develop AI, but they make critical decisions about its implementation within their organization.
Target_Population	policymakers	Policy Shapers: This code is targeted to those in positions to influence policy and make significant decisions about the use and regulation of AI, including lawmakers and regulators.
Target_Population	population_other	Other: This code refers to studies that focus on a definite, identifiable group discussed in the publication distinct from the other options. Examples could include women, judges, doctors, or any other specific demographic or professional group.
Target_Population	public	General Public Subjects of AI: This refers to studies targeted towards the broader population, not specifically individuals in academic or professional training. The focus here is on AI literacy for the everyday citizen.
Target_Population	STEM_workforce	(industry) AI Deployers: This category includes those who apply or implement AI tools in practical settings. They might not be AI experts but utilize AI technology fairly directly in their work.

(*continued on next page*)

Table A.1 (*continued*)

Code Family	Code	Code Description or Guidance
Target_Population	students	Discipline or non-discipline specific students. Use this code if the level of student is not specified and the code graduate, postsecondary, primary, primary, secondary cannot apply.
Target_Population	students graduate	If target population is students, this specifies they are graduate school students.
Target_Population	students postsecondary	If target population is students, this specifies they are postsecondary school students. Postsecondary can include universities and colleges, also trade and vocational schools.
Target_Population	students primary	If target population is students, this specifies they are primary school students.
Target_Population	students secondary	If target population is students, this specifies they are secondary school students. Typically what is high school in the US.
Target_Population	teachers	Targeting either K-12 or Higher Ed teachers who will be teaching about AI ethics
Ethics_Learning_Goal	communication_confidence	Character transformation. The aim is to enhance communication skills, boost confidence, and voice personal aspects of moral concerns. This involves fostering the ability to articulate ethical viewpoints confidently and commit to ethical decisions. It's important in ethics education as it empowers students to actively participate in ethical discussions and stand up for their beliefs. Communication, commitment, voice and confidence.
Ethics_Learning_Goal	EthicalReasoning	This goal focuses on students' ethical reasoning or moral reasoning capacities. Ethical reasoning involves the ability to make decisions and judgements about ethical situations. Typically involves awareness, judgment, evaluation, and decision. Ethical reasoning might include considering stakeholders, their values, pros and cons of their values, what it looks like it discipline-specific context, and making decisions.
Ethics_Learning_Goal	experience_intuition	Personal morality, intuitions, emotions, deep connecting parts of the individual. The goal is to immerse students in experiences and connect their personal identities to professional ethics. Personal experience or intuition for ethics. This develops a priori reasoning for ethics. This involves exposing learners to real-world ethical scenarios and helping them understand their intuitive and emotional responses to scenarios. This is crucial in ethics education as it allows students to connect theory with practice.
Ethics_Learning_Goal	knowledge_theory	Foundational knowledge. The goal is to acquire knowledge about theories, philosophies, codes, or laws related to ethics. This involves learning the established frameworks and principles that guide ethical behavior. This is essential in ethics education as it provides the theoretical foundation for ethical understanding and action. Knowledge, theory, or philosophy.
Ethics_Learning_Goal	other	If the learning goal cannot be reasonably fit into one of the other learning goal categories
Educational_Context	active	Active (vs Passive) learning engages students in the learning process, requiring them to actively process and apply the information.
Educational_Context	collaborative	Collaborative learning (vs self-directed) involves a group of individuals working together towards the same learning objectives.
Educational_Context	passive	Passive (vs active) learning involves students being exposed to information without active engagement or processing.
Educational_Context	self	Self-directed learning (vs collaborative) involves an individual taking the initiative and being responsible for their learning experience.
Ethics_Topics	autonomy	Human freedom and power to choose/decide what to do on their own. Typically discussed in the context of an AI reducing human autonomy. The more an AI makes decisions for a human, the less autonomous they are.
Ethics_Topics	bias	Bias (including algorithmic bias): The tendency of a system to favor one group of people over another.
Ethics_Topics	deception	Honesty/Deceptive Work or Business Practices (e.g., bribes): The importance of being honest and transparent in business dealings.
Ethics_Topics	diversity_equity_inclusion	Diversity, Equity, and/or Inclusion: The need to ensure that all people are represented and treated fairly in the technology industry.
Ethics_Topics	ethical_theories	Ethical Theories (e.g., Utilitarianism or Ethics of Care): A system of moral principles that can be used to guide ethical behavior in technology.
Ethics_Topics	ethics_codes	Codes of Ethics: A set of principles that guide ethical behavior in technology typically from a specific organization or institution. Something like ACM code of ethics of IEEE code of ethics.
Ethics_Topics	inequality_power_fairness	Inequality/Power Imbalances/Fairness: The unequal distribution of power and resources in society.
Ethics_Topics	intellectual_property	Intellectual Property (IP): The legal rights that protect creative works and inventions.
Ethics_Topics	misinformation	Disinformation/Misinformation: The spread of false or misleading information.
Ethics_Topics	other	Other (please describe): Any other ethical considerations that were not mentioned above.
Ethics_Topics	policing	The use of government resources for surveillance of citizens or enforcement of policy via AI technology.
Ethics_Topics	privacy	Privacy/Surveillance: The collection and use of personal data by governments and corporations.
Ethics_Topics	risk_safety	Risk/Safety: Risk of AI to be used for the potential for harm or danger.
Ethics_Topics	robotics	Robotics: AI's integration with the field of robotics and the use cases of robotics in society, work, or daily life.
Ethics_Topics	social_good_computing	Humanitarian Computing/Computing and Social Good: The use of technology to solve social problems.
Ethics_Topics	social_justice	Social Justice: The pursuit of justice for all people, regardless of their race, ethnicity, gender, sexual orientation, or other social group.
Ethics_Topics	sustainability	Environmental Ethics/Sustainability: The ethical principles that guide our relationship with the environment.
Ethics_Topics	tech_access	Access to Technology (e.g., the Digital Divide): The gap between those who have access to technology and those who do not.
Ethics_Topics	technological_design	The approach to the technical design of ethics within technology. The technical approaches to embedding ethics in AI
Ethics_Topics	transparency_explainability	Explainability/Transparency: Technical construct about the system ability to understand how a system works and why it makes the decisions it does.
Ethics_Topics	universal_design	Universal Design/Design for Disabilities/Accessibility: The design of products and services that are accessible to people with disabilities.
Pedagogical_Delivery	Hybrid	The educational material is delivered intentionally with both in person and online components. Intentionally designed.
Pedagogical_Delivery	In-person	The educational material is delivered in person. Whether the design of the intervention/course/module was meant for in-person.
Pedagogical_Delivery	Online	The educational material is delivered online.
Pedagogy	Application of theory, codes, laws, rules	Reviewing codes or theories to learn ethics.
Pedagogy	Case studies	Utilizing case studies as a way to teach ethics.
Pedagogy	Discussion	Discussion, debate, verbal presentation, peer workshopping, brainstorming, role-playing.

(*continued on next page*)

Table A.1 (*continued*)

Code Family	Code	Code Description or Guidance
Pedagogy	Experimentation	Experimenting with ethics tools, heuristics, processes, decision-making frameworks. Creating, experimenting, reasoning through ethical dilemmas Creating, experimenting, reasoning.
Pedagogy	Group_project	Group based work to create some sort of deliverable. Group project work.
Pedagogy	HandsOn	Engaging students in active and practical tasks that either physically or abstractly works with AI systems and/or ethical frameworks to learn about AI ethics topics. Could be programming, design, and development tasks or abstractly applying frameworks to cases.
Pedagogy	Lecture	Intentional class dedicated toward a lecture to learn about topics.
Pedagogy	Non-traditional	When the instructor intentionally tries to incorporate non-traditional methods for teaching. Could include playing a game to learn about ethics or ethical dilemmas, watching a sci-fi movie, reading a sci-fi book, reading or creating comics/humor to engage with the ethics topic.˙
Pedagogy	other	Other pedagogy that cannot be easily categorized.
Pedagogy	Real-world exposure	Using real-world exposure, project-based learning. Real-world exposure, project-based learning to teach ethics. Typically out of the classroom (though maybe a real company will come into the classroom). For example, engineering students go to a local manufacturing firm and gain experience on the shop floor.
Pedagogy	Written	Employing paper, essay, reflection paper, essay, reflection as a means to learn ethics.˙
Challenges	Challenges	Challenges to teaching ethics. Code mentions of why it is difficult to teach, instruct, or deliver ethics. Inductively code. Highlight segment/unit of analysis where the challenge is listed and code challenge.
Quantitative_Assessment	Comparative	Experimental pre/post-test Experimental pre/post-test.
Quantitative_Assessment	Exam	Quiz, test, exam Quiz, test, exam.
Quantitative_Assessment	Instrument	Using established instruments such as ESIT, DIT2 ESIT, DIT2.
Quantitative_Assessment	Rubrics	Quantifying qualitative data such as rubrics or coding responses or perceptions. Rubrics, quantified codes Rubrics, quantified codes.
Quantitative_Assessment	Surveys	Deploying a quantitative survey to students gain insight into constructs. Not an experiment pre/post-test design, but just survey instrument given to students.
Qualitative_Assessment	Artifact assessment	Analyzing something that was created by the subjects Homework assignment, learning activity.
Qualitative_Assessment	Interviews	Intentional method to collect qualitative data from human subjects. Interview, focus groups Interview, focus groups.
Qualitative_Assessment	Observation	Observations of learning or behaviors.
Qualitative_Assessment	Surveys	Open ended surveys; course evaluations, student experience reflections, perceptions course evaluation, student experience reflections, perceptions.
Qualitative_Assessment	Written	Intentional method to collect qualitative data from human subjects. Interview, focus groups Interview, focus groups.

Appendix B. Detailed PRISMA Documentation

B.1 Title

AI Ethics Education: A Systematic Literature Review.

B.2 Background

Education for AI ethics is necessary for future or current professionals to navigate the challenges posed by ubiquitous AI integrations in the workplace, society, or daily life. The field of AI ethics education promotes the responsible use and development of AI by equipping individuals–-students, professionals, and the public–with the knowledge or skills needed to act ethically alongside life with AI. However, it is unclear which educational efforts are most effective for this grand goal. In this study, we take a look at the focused efforts on AI ethics education that occurred prior to the arrival of commercial large language models like ChatGPT to best understand the educational principles from the foundation of the field to infer the future trajectory of the AI ethics education.

B.3 Research Questions

ID	Research Question	Motivation
RQ1	What do the early years of (formalized) AI ethics education interventions, from 2018 to 2023, suggest about the state and future trajectory of the field?	To state the intention of the study and scope the sample and implications of research.
RQ1.1	To what extent are content, assessment, and pedagogy aligned for AI ethics education interventions in the early years of AI ethics education?	To provide operational guidance about how the primary RQ will be answered through educational alignment.
RQ2	To what extent does AI ethics education adopt or diverge from best practices in computing and engineering ethics education?	To identify strengths and weaknesses of the field, in a nascent state, compared to more established fields

B.4 Search Strategy and Study Selection

Three conceptual constructs are under investigation in this study: AI, ethics, and education. The search terms were generated from the first two constructs and the third construct was captured through the manual screening process.

Key Term	Search Terms	Rationale
AI Ethics	AI Ethics, Artificial Intelligence Ethics, Responsible AI, Responsible Artificial Intelligence, Ethical AI, Ethical Artificial Intelligence	The authors experimented with many formulations of the search terms and search strings, and often experienced unreliable and unwieldy results from searching "AI" (with synonyms) and "ethics" (with synonyms). Papers would frequently use both terms as a throw-away focus within their paper, rather than a centralized focus on "AI ethics" as a field of its own. Thus, we forced a joint term of "AI ethics" and "responsible AI" in the literature to find focused papers in the substantive field of AI ethics and educational intervention within such.
Education	No search terms. Manually screened.	Our study was not limited to educational interventions in traditional classroom or 'Education' formats. We were interested in professional training and development, extracurricular workshops for students and professionals, public literacy efforts, and workforce reskilling efforts as well as the traditional K12 and Higher Education interventions for AI ethics. For comprehensiveness, we chose to manually screen for 'education' since a concise set of keywords could not capture the wide-ranging attempts across these fields.

Eligible records had to be published within 2018–2023, written in English, and published in conference proceedings or journal venues. The research team made note *not* to restrict the search to only top journals or conference venues to have a more inclusive and broader array of collected records. The research team identified the following databases for searching records: Web of Science, Scopus, PhilPapers, ACM Digital Library, and IEEE Xplore. These databases were chosen due to comprehensiveness over the substantive "AI ethics" field. Each of these databases required unique search syntax that modified our primary search string. The search was restricted to January 1st, 2018, to December 31st, 2023, and each search was conducted on February 16th, 2023. See below for each unique search string and the number of records exported from each search.

Database	Search String	Count
Web of Science	(((((AB=(("AI Ethics' OR "Artificial Intelligence Ethics' OR "responsible AI' OR "responsible artificial intelligence' OR "ethical AI' OR "ethical artificial intelligence"))) OR TI=(("AI Ethics' OR "Artificial Intelligence Ethics' OR "responsible AI' OR "responsible artificial intelligence' OR "ethical AI' OR "ethical artificial intelligence"))) OR AK=(("AI Ethics' OR "Artificial Intelligence Ethics' OR "responsible AI' OR "responsible artificial intelligence' OR "ethical AI' OR "ethical artificial intelligence"))) AND PY=(2018–2023)) AND LA=(English)) AND DT=(Article OR Proceedings Paper)	563
Scopus	TITLE-ABS-KEY (("AI Ethics' OR "Artificial Intelligence Ethics' OR "responsible AI' OR "responsible artificial intelligence' OR "ethical AI' OR "ethical artificial intelligence")) AND PUBYEAR >2018 AND DOCTYPE (ar OR cp)	838
PhilPapers *	(AI Ethics | Artificial Intelligence Ethics | responsible AI | responsible artificial intelligence | ethical AI | ethical artificial intelligence) **This database's search is limited; All fields, years, languages, and document types were searched for and collected. Then, the inclusion-exlclusion criteria were filtered out as part of screening process.*	792
ACM Digital Library *	[Title: "ai ethics"] OR [Title: "artificial intelligence ethics"] OR [Title: "responsible ai"] OR [Title: "responsible artificial intelligence"] OR [Title: "ethical ai"] OR [Title: "ethical artificial intelligence"] OR [Abstract: "ai ethics"] OR [Abstract: "artificial intelligence ethics"] OR [Abstract: "responsible ai"] OR [Abstract: "responsible artificial intelligence"] OR [Abstract: "ethical ai"] OR [Abstract: "ethical artificial intelligence"] OR [Keywords: "ai ethics"] OR [Keywords: "artificial intelligence ethics"] OR [Keywords: "responsible ai"] OR [Keywords: "responsible artificial intelligence"] OR [Keywords: "ethical ai"] OR [Keywords: "ethical artificial intelligence"] AND [E-Publication Date: (January 01, 2018 TO 12/31/2023)] **This database was not filtered for language or document type during the search phase. These criteria were filtered during the screening process.*	221
IEEE Xplore *	("AI Ethics' OR "Artificial Intelligence Ethics' OR "responsible AI' OR "responsible artificial intelligence' OR "ethical AI' OR "ethical artificial intelligence") **Dates were filtered on the website graphical user interface: 2018–2023. Other criteria were not able to reliably be filtered so all document types, languages, and fields were searched. These criteria were filtered during the screening process.*	254

The completed search yielded a total of 2668 records (before removing duplicates). We exported all records from each database in BibTeX format to facilitate standardized record keeping, remove duplicates, and manage metadata for each record. We utilized the Zotero reference manager's automated duplicate detection to remove duplicate records.

Conceptual inclusion and exclusion criteria were, broadly, construed on the focus of whether AI, Ethics, and Education are in substantive focus of the paper, and whether there was an empirical review of an educational intervention to achieve AI ethics. To determine our final set of papers we used a multi-phase exclusion procedure.

Phase	Description	Records left
1	Starting $n = 2668$. We excluded duplicate records and non-journal or non-conference records that were missed in the databases' export functionality. We utilized Zotero's automated item-type detection and duplicate-record detection to remove a total of 783 duplicate records and 318 non-journal and non-conference records (such as theses, public reports, or books).	1567
2	We screened for the third construct under investigation: **education**. First, we ensured each record in our corpus had an abstract. Then, we exported records from Zotero to CSV where we used Excel to read and process each abstract. We used an Excel macro script to highlight relevant keywords (below) to facilitate manually reading each abstract. Records that did not have an education keyword were marked for removal. Once each abstract was processed, records This resulted in a total of 886 records removed, leaving records in at least partial relevance to AI, ethics, and education.	681
3	The primary researcher read each abstract, located the highlighted keyword, and determined whether the education keyword was used in substantive focus for the scope of our research questions. Instances out of scope might include 'learning' being falsely included by 'machine learning' or 'training' being discussed as 'data training' rather than training efforts for on AI ethics. 515 records were removed.	166
4	We manually screened the full-text to make a final decision about whether the substantive focus of each paper was about AI ethics education. To start, the team immediately identified 11 full-text records that ended up being the wrong item type, not in English, or had no full text available. Then, the research team appraised the records for substantive discussions of AI, ethics, and education. For example, the phrase "AI Ethics" may have been used in the manner of an author's secondary remark, like "AI, as well as AI ethics, is an important discussion." −73 removed due to education not being a substantive focus	56

(*continued on next page*)

(*continued*)

Phase	Description	Records left
	−14 removed due to ethics not being within scope −12 removed due to a focus of AI for education −11 removed due to metadata eligibility criteria.	
5	We removed purely theoretical or conceptual papers in this stage. We were interested in AI ethics education as it is practiced in the early years of the field. To do this, we were concerned with empirical investigations of times when an educational intervention on AI ethics was delivered. To answer our research questions, we needed empirical detail about the educational practices of AI ethics to determine its educational alignment and infer the state of the field moving forward. We removed 31 theoretically-oriented papers. Theoretically-oriented papers were about how educational strategies and interventions incorporate AI ethics, rather than being specific interventions themselves. Inversely, empirically-oriented intervention papers either detail educational course designs, workshops, scalable modules for AI ethics education, or empirical reviews of interventions of AI ethics conducted from other authors. In result, we systematically deduced a dataset of papers on AI ethics education interventions from the 'early years' of the field.	25*

* Each record included (Bae et al., 2022; Bendechache et al., 2021; Chklovski et al., 2020, pp. 34–35; Eguchi et al., 2021; Forsyth et al., 2021, pp. 1–2; Garrett et al., 2020, pp. 272–278; Gong et al., 2020; N. Green, 2021, pp. 519–524; N. L. Green & Crotts, 2020; Hishiyama & Shao, 2022; Hod et al., 2022; Javed et al., 2022; Kim et al., 2022; Ng et al., 2022; Ottenbreit-Leftwich et al., 2022; Perchik et al., 2023; Shih et al., 2021; Stoyanovich, 2022; Taylor & Deb, 2021; Tuovinen & Rohunen, 2021; Van Brummelen et al., 2021; Williams et al., 2022; Zhang et al., 2023; Zhao et al., 2022; Zhou et al., 2022).

Following PRISMA reporting, the following figure (Figure B.1) is the official technical figure for the PRISMA procedure(Page et al., 2021). See above for details on each step.

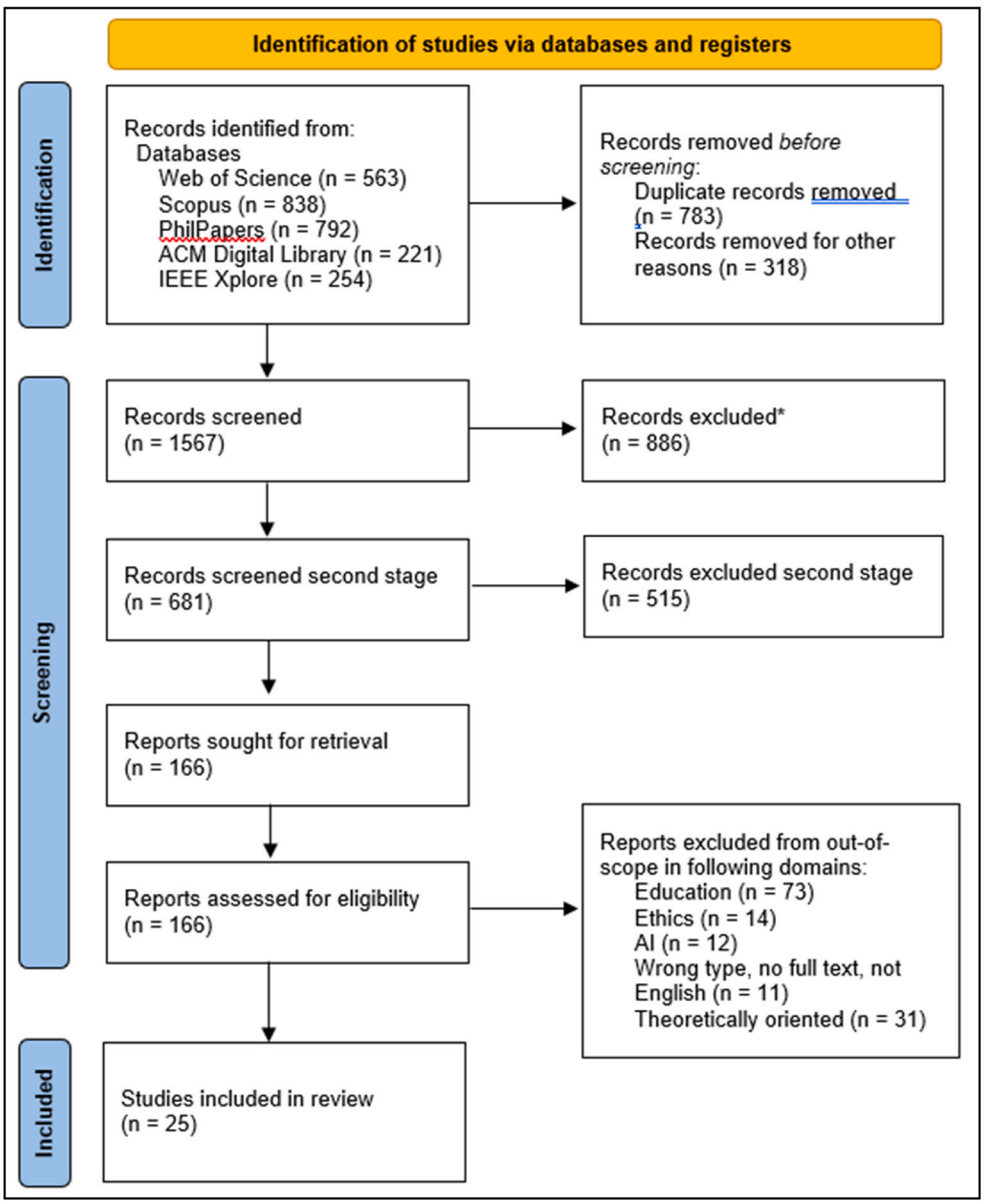

Figure B.1Official PRISMA Flow Diagram for current study.
*Assisted by automation tools. Used script to mark records without any 'education' related words.

Education keywords for abstract screening

k12	*vocation*	*school*
education	*teach*	*curricul*
pedagog	*learn*	*instruction*
train	*workplace development*	*reskill*
human performance improvement	*professional growth*	*litera*
workforce development	*upskill*	*awareness*
scholarship	*competenc*	

B.5 Codebook Development

In the main body of this paper we reported on the codebook families, a description, and associated codes for each family in Table 1. We listed the full operational codebook used to extract data from our paper dataset in Appendix A. Here, we will give an overview of the development process for this codebook.

We used this codebook to extract salient data from our dataset of papers on AI ethics education. Under a deductive approach, we identified pre-existing meaningful classifications of concepts that could be used to answer our research questions.

We primarily leveraged the principle of *backwards design* via the Content, Assessment, and Pedagogy framework to extract data about the (1) content of AI ethics education, (2) pedagogy used to implement AI ethics education, and (3) assessment mechanisms used to support student learning progressions for AI ethics education. Then, we included additional code families to extract data on the authorship, research design, and context of the papers.

Despite being clear in these high-level distinctions, we had to go through multiple iterations of the codebook to ensure **codebook-data fit**. While screening the records in Phases 3, 4, and 5, we constantly compared the dataset to our codebook and conducted trial runs to simulate what data analysis might look like under certain inclusion/exclusion criteria. For instance, while manually appraising the full-text records in Phase 4, the researchers would each take a random subsample of papers and code the papers as if it were the final dataset. As a result, we could identify what our codebook was able to capture and what it might have missed. With this information, (1) we increasingly refined our codebook, (2) familiarized ourselves with the data, and (3) increased clarity over our inclusion/exclusion criteria and research scope.

Before moving to the final coding procedure for data extraction over our final dataset, we agreed on a final codebook that was not to change any further. This is what is reported in Appendix A. This codebook allows us to standardize the data collection from the heterogeneous dataset that included papers of different lengths, types, and structures.

B.6 Interrater Reliability

Three members of the research team used the codebook to code each of the 25 papers in our dataset. Each week, each member was tasked with coding two to four papers. At the end of the week, the primary researcher compiled each of the coded papers and exported a document-code matrix to compare agreements and disagreements between each author. It is important to note here that these were *document-level* codes. I.e., if Author 1 and Author 2 both coded a paper to concern "Privacy," that was considered an apt level of agreement *without needing* the authors to code "Privacy" in the same exact sentence of the paper. This document-code matrix was used to examine levels of agreement within each code, each code family, and each document between each of the three research analysts. Then, in weekly meetings, the research team met to discuss discrepancies in the codes, identify areas of conceptual ambiguity, and reach agreement on the set of codes applied to each paper. Once agreement was met, each researcher moved on to the next set of papers. Thus, this intercoder agreement (ICA) process was used to (1) facilitate weekly team meetings to ensure a standard and systematic process for this project and (2) determine final data extraction for each paper in the dataset.

The resulting set of codes applied to each paper concluded the data extraction for this project. We then conducted qualitative techniques to analyze the data: we reviewed the substance of each code, organized codes into categories, and rectified findings and themes from the code. For data analysis, refer to the main body of the paper.

References

AI Singapore. (2024). AI Singapore. https://aisingapore.org/.

AI4K12. (2020). *AI4K12—sparking Curiosity in AI*. AI4K12. https://ai4k12.org/.

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics, 2*(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7

Albarracín, D., Fayaz-Farkhad, B., & Granados Samayoa, J. A. (2024). Determinants of behaviour and their efficacy as targets of behavioural change interventions. *Nat. Rev. Psychology, 3*(6), 377–392. https://doi.org/10.1038/s44159-024-00305-0

Arnold, Z., Schiff, D. S., Schiff, K. J., Love, B., Melot, J., Singh, N., Jenkins, L., Lin, A., Pilz, K., Enweareazu, O., & Girard, T. (2024). Introducing the AI governance and regulatory archive (AGORA): An analytic infrastructure for navigating the emerging AI governance landscape. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 7*, 39–48. https://doi.org/10.1609/aies.v7i1.31615

ATLAS.ti. (2023). ATLAS.ti qualitative data analysis software. *ATLAS.ti Scientific Software Development GmbH* [Computer software] Version 23.4.0. https://atlasti.com.

Bae, J., Lee, J., & Cho, J. (2022). Analysis of AI ethical competence to computational thinking. *International. J. Inf. Visualization, 6*(2–2), 506. https://doi.org/10.30630/joiv.6.2-2.1126

Barab, S. (2014). Design-based research: A methodological toolkit for engineering change. In K. Sawyer (Ed.), *The cambridge handbook of the learning sciences* (2nd ed., pp. 151–170). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.

Bargh, J. A. (2022). The cognitive unconscious in everyday life. In A. S. Reber, & R. Allen (Eds.), *The cognitive unconscious* (1st ed., pp. 89–112). York: Oxford University PressNew. https://doi.org/10.1093/oso/9780197501573.003.0005.

Barkhuff, G., Borenstein, J., Schiff, D., Uchidiuno, J., & Zegura, E. (2024). Considerations for improving comprehensive undergraduate computing ethics education. *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V, 2*, 1560–1561. https://doi.org/10.1145/3626253.3635557

Bendechache, M., Tal, I., Wall, P., Grehan, L., Clarke, E., Odriscoll, A., Der Haegen, L. V., Leong, B., Kearns, A., & Brennan, R. (2021). AI in my life: AI, ethics & privacy workshops for 15-16-year-olds. *13th ACM Web science conference 2021* (pp. 34–39). https://doi.org/10.1145/3462741.3466664

Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability, 21*(1), 33–46. https://doi.org/10.1007/s11092-008-9064-9

Bingham, A., & Witkowsky, P. (2022). Deductive and inductive approaches to qualitative data analysis. In C. Vanover, P. Mihas, & J. Saldaña (Eds.), *Analyzing and interpreting qualitative research: After the interview* (1st ed.). SAGE Publications, Inc.

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

Borrego, M., Foster, M. J., & Froyd, J. E. (2014). Systematic literature reviews in engineering education and other developing interdisciplinary fields. *Journal of Engineering Education, 103*(1), 45–76. https://doi.org/10.1002/jee.20038

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *Research designs: Quantitative,* *qualitative, neuropsychological, and biological: Vol. 2. APA handbook of research methods in psychology* (pp. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004.

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), 1–36. https://doi.org/10.1145/3634685

Burton, E., Goldsmith, J., & Mattei, N. (2018). How to teach computer ethics through science fiction. *Communications of the ACM, 61*(8), 54–64. https://doi.org/10.1145/3154485

California, S. of (2024). *California, NVIDIA launch first-of-its-kind AI collaboration*. Governor of California. https://www.gov.ca.gov/2024/08/09/california-nvidia-launch-first-of-its-kind-ai-collaboration/.

Cengage. (2024). *2024 graduate employability report: Preparing students for the GenAI-driven workplace*. Cengage group. https://cengage.widen.net/s/bmjxxjx9mm/cg-2024-employability-survey-report.

Chen, Y.-C., & Ahn, M. (2022). Governing AI systems for public values: Design principles and a process framework. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The oxford handbook of AI governance* (1st ed., pp. 421–440). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.31.

Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Comput. Educ.Open, 6*, Article 100171. https://doi.org/10.1016/j.caeo.2024.100171

Chiu, T. K. F., Meng, H., Chai, C.-S., King, I., Wong, S., & Yam, Y. (2022). Creation and evaluation of a pretertiary artificial intelligence (AI) curriculum. *IEEE Transactions on Education, 65*(1), 30–39. https://doi.org/10.1109/TE.2021.3085878

Chklovski, T., Jung, R., & Young, K. (2020). Engaging families in a 10-week, AI global competition. *Proceedings of the 2nd ACM SIGSOFT international workshop on education through advanced software engineering and artificial intelligence*. https://doi.org/10.1145/3412453.3423199

Clancy, R. F., & Zhu, Q. (2023). Why should ethical behaviors Be the ultimate goal of engineering ethics education? *Business & Professional Ethics Journal, 42*(1), 33–53. https://doi.org/10.5840/bpej202346136

Cohen, A. (1987). Instructional alignment: Searching for a magic bullet. *Educational Researcher, 16*(8), 16–20. https://doi.org/10.3102/0013189X016008016

Corrêa, N. K., Galvão, C., Santos, J. W., Pino, C. D., Pinto, E. P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & Oliveira, N. de (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns, 4*(10). https://doi.org/10.1016/j.patter.2023.100857

Digital Education Council. (2024). Digital education council global AI student survey 2024. *Digital education council*. https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024.

Dai, Y., Liu, A., Qin, J., Guo, Y., Jong, M. S.-Y., Chai, C.-S., & Lin, Z. (2023). Collaborative construction of artificial intelligence curriculum in primary schools. *Journal of Engineering Education, 112*(1), 23–42. https://doi.org/10.1002/jee.20503

Dempsey, M., McBride, K., Haataja, M., & Bryson, J. (2022). Transnational digital governance and its impact on artificial intelligence. In *The oxford handbook of AI governance*. Oxford University Press.

DeVon, C. (2024). Google, Adobe and IBM are aiming to help millions gain AI skills—here's what to know. *CNBC*. https://www.cnbc.com/2024/10/25/google-adobe-and-ibm-aiming-to-help-millions-gain-ai-skills.html.

Dieterle, E., Dede, C., & Walker, M. (2022). *The cyclical ethical effects of using artificial intelligence in education*. AI & SOCIETY. https://doi.org/10.1007/s00146-022-01497-w

Dorie, B. L., Dankenbring, C. A., Denick, D. L., Ferguson, D., Huff, J., Phillips, C., Schimpf, C., & Cardella, M. E. (2012). File: A taxonomy of formal and informal learning environments. *2012 ASEE annual conference & exposition. 2012 ASEE annual conference & exposition, valparaiso, Indiana*.

Doroudi, S. (2022). The intertwined histories of artificial intelligence and education. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-022-00313-2

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eguchi, A., Okada, H., & Muto, Y. (2021). Contextualizing AI education for K-12 students to enhance their learning of AI literacy through culturally responsive approaches. *KI - Kunstliche Intelligenz, 35*(2), 153–161. https://doi.org/10.1007/s13218-021-00737-3

European Commission. (2021). Finland AI strategy report. https://ai-watch.ec.europa.eu/countries/finland/finland-ai-strategy-report_en.

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. http://data.europa.eu/eli/reg/2024/1689/oj/eng.

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. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3518482

Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy Technology, 32*(2), 185–193. https://doi.org/10.1007/s13347-019-00354-x

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

Forsyth, S., Dalton, B., Foster, E. H., Walsh, B., Smilack, J., & Yeh, T. (2021). Imagine a more ethical AI: Using stories to develop teens' awareness and understanding of artificial intelligence and its societal impacts. *2021 conference on research in equitable and sustained participation in engineering, computing, and technology (RESPECT)*. https://doi.org/10.1109/RESPECT51740.2021.9620549

Gambelin, O. (2021). Brave: What it means to be an AI Ethicist. *AI and Ethics, 1*(1), 87–91. https://doi.org/10.1007/s43681-020-00020-5

Garrett, N., Beard, N., & Fiesler, C. (2020). More than "if time allows": The role of ethics in AI education. *Proceedings of the AAAI/ACM conference on AI*. Ethics, and Society. https://doi.org/10.1145/3375627.3375868

Gong, X., Tang, Y., Liu, X., Jing, S., Cui, W., Liang, J., & Wang, F.-Y. (2020). K-9 artificial intelligence education in qingdao: Issues, challenges and suggestions. *2020 IEEE international conference on networking, sensing and control, ICNSC 2020*. https://doi.org/10.1109/ICNSC48988.2020.9238087

Google. (2024). *How we're improving AI literacy in young people*. Google. https://blog.google/technology/families/improving-ai-literacy-in-young-people/.

Graneheim, U. H., Lindgren, B.-M., & Lundman, B. (2017). Methodological challenges in qualitative content analysis: A discussion paper. *Nurse Education Today, 56*, 29–34. https://doi.org/10.1016/j.nedt.2017.06.002

Green, N. (2021). An AI ethics course highlighting explicit ethical agents. *Proceedings of the 2021 AAAI/ACM conference on AI*. Ethics, and Society. https://doi.org/10.1145/3461702.3462552

Green, N. L., & Crotts, L. J. (2020). Argument schemes for AI ethics education. *Computational models of natural argument (CMNA 2020)* (pp. 41–50).

Griffin, T. A., Green, B. P., & Welie, J. V. M. (2024). The ethical wisdom of AI developers. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00458-x

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

Hagendorff, T. (2022a). A virtue-based framework to support putting AI ethics into practice. *Philosophy Technology, 35*(3), 55. https://doi.org/10.1007/s13347-022-00553-z

Hagendorff, T. (2022b). Blind spots in AI ethics. *AI and Ethics, 2*(4), 851–867. https://doi.org/10.1007/s43681-021-00122-8

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834.

Hartikainen, S., Rintala, H., Pylväs, L., & Nokelainen, P. (2019). The concept of active learning and the measurement of learning outcomes: A review of research in engineering higher education. *Education Sciences, 9*(4). https://doi.org/10.3390/educsci9040276. Article 4.

Hess, J. L., & Fore, G. (2017). A systematic literature review of US engineering ethics interventions. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-017-9910-6

Hishiyama, R., & Shao, T. (2022). Educational effects of the case method in teaching AI ethicsA. Rocha, H. Adeli, G. Dzemyda, & F. Moreira (Eds.). *Lecture Notes Network. Syst., 468*, 226–236. https://doi.org/10.1007/978-3-031-04826-5_22

Hod, S., Chagal-Feferkorn, K., Elkin-Koren, N., & Gal, A. (2022). Data science meets law. *Communications of the ACM, 65*(2), 35–39. https://doi.org/10.1145/3506575

Holmes, W., & Porayska-Pomsta, K. (2022). *The ethics of artificial intelligence in education: Practices, challenges, and debates* (1st ed.). Routledge. https://doi.org/10.4324/9780429329067

Howley, I., Darakhshan, M., & Peck, E. (2022). Integrating AI ethics across the computing curriculum. In W. Holmes, & K. Porayska-Pomsta (Eds.), *The ethics of artificial intelligence in education: Practices, challenges, and debates* (1st ed., pp. 255–270). Routledge. https://doi.org/10.4324/9780429329067.

H.R.6791 - 118th Congress (2023-2024). (2023). Artificial intelligence literacy act of 2023. https://www.congress.gov/bill/118th-congress/house-bill/6791.

Javed, R. T., Nasir, O., Borit, M., Vanhée, L., Zea, E., Gupta, S., Vinuesa, R., & Qadir, J. (2022). Get out of the BAG! Silos in AI ethics education: Unsupervised topic modeling analysis of global AI curricula. *Journal of Artificial Intelligence Research, 73*, 933–965. https://doi.org/10.1613/jair.1.13550

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

Karabulut-Ilgu, A., Jaramillo Cherrez, N., & Jahren, C. T. (2018). A systematic review of research on the flipped learning method in engineering education. *British Journal of Educational Technology, 49*(3), 398–411. https://doi.org/10.1111/bjet.12548

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns, 2*(9), Article 100314. https://doi.org/10.1016/j.patter.2021.100314

Khan, A. A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., Fahmideh, M., Niazi, M., & Akbar, M. A. (2022). Ethics of AI: A systematic literature review of principles and challenges. *Proceedings of the international conference on evaluation and assessment in software engineering*. https://doi.org/10.1145/3530019.3531329

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies, 27*(5), 6069–6104. https://doi.org/10.1007/s10639-021-10831-6

Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving Schools, 19*(3), 267–277. https://doi.org/10.1177/1365480216659733

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*, 633–705.

Kuipers, B. (2020). Perspectives on ethics of AI: Computer science. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The oxford handbook of ethics of AI* (pp. 419–441). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.27.

Lapsley, D. K., & Hill, P. L. (2008). On dual processing and heuristic approaches to moral cognition. *Journal of Moral Education, 37*(3), 313–332. https://doi.org/10.1080/03057240802227486

Lauer, D. (2021). You cannot have AI ethics without ethics. *AI and Ethics, 1*(1), 21–25. https://doi.org/10.1007/s43681-020-00013-4

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI conference on human factors in computing systems*. https://doi.org/10.1145/3313831.3376727

Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2024). Responsible AI pattern catalogue: A collection of best practices for AI governance and engineering. *ACM Computing Surveys, 56*(7), 1–35. https://doi.org/10.1145/3626234

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

Lyon, J. A., & Magana, J. A. (2020). Computational thinking in higher education: A review of the literature. *Computer Applications in Engineering Education, 28*(5), 1174–1189. https://doi.org/10.1002/cae.22295

Magana, A. J. (2022). The role of frameworks in engineering education research. *Journal of Engineering Education, 111*(1), 9–13. https://doi.org/10.1002/jee.20443

Martin, F. (2011). Instructional design and the importance of instructional alignment. *Community College Journal of Research and Practice, 35*(12), 955–972. https://doi.org/10.1080/10668920802466483

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics, 160*(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3

Martin, D. A., Conlon, E., & Bowe, B. (2021). A multi-level review of engineering ethics education: Towards a socio-technical orientation of engineering education for ethics. *Science and Engineering Ethics, 27*(5), 60. https://doi.org/10.1007/s11948-021-00333-6

MIT. (2024). RAISE initiative: Responsible AI for social empowerment and education. https://raise.mit.edu/.

Mitcham, C., & Englehardt, E. E. (2019). Ethics across the curriculum: Prospects for broader (and deeper) teaching and learning in research and engineering ethics. *Science and Engineering Ethics, 25*(6), 1735–1762. https://doi.org/10.1007/s11948-016-9797-7

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a service: A pragmatic operationalisation of AI ethics. *Minds and Machines, 31*(2), 239–256. https://doi.org/10.1007/s11023-021-09563-w

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

Munn, L. (2023). The uselessness of AI ethics. *AI and Ethics, 3*(3), 869–877. https://doi.org/10.1007/s43681-022-00209-w

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers & Education: Artificial Intelligence, 2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ng, D. T. K., Luo, W., Chan, H. M. Y., & Chu, S. K. W. (2022). Using digital story writing as a pedagogy to develop AI literacy among primary students. *Computers & Education: Artificial Intelligence, 3*, Article 100054. https://doi.org/10.1016/j.caeai.2022.100054

Nguyen, D., & Hekman, E. (2024). The news framing of artificial intelligence: A critical exploration of how media discourses make sense of automation. *AI & Society, 39*(2), 437–451. https://doi.org/10.1007/s00146-022-01511-1

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies, 28*(4), 4221–4241. https://doi.org/10.1007/s10639-022-11316-w

Ottenbreit-Leftwich, A., Glazewski, K., Jeon, M., Jantaraweragul, K., Hmelo-Silver, C. E., Scribner, A., Lee, S., Mott, B., & Lester, J. (2022). Lessons learned for AI education with elementary students and teachers. In *International journal of artificial intelligence in education*. SPRINGER. https://doi.org/10.1007/s40593-022-00304-3.

Padiyath, A. (2024). A realist review of undergraduate student attitudes towards ethical interventions in technical computing courses. *ACM Transactions on Computing Education, 24*(2), 1–19. https://doi.org/10.1145/3639572

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews, 10*(1), 89. https://doi.org/10.1186/s13643-021-01626-4

Pant, A., Hoda, R., Spiegler, S. V., Tantithamthavorn, C., & Turhan, B. (2024). Ethics in the age of AI: An analysis of AI practitioners' awareness and challenges. *ACM Transactions on Software Engineering and Methodology, 33*(3), 1–35. https://doi.org/10.1145/3635715

Perchik, J. D., Smith, A. D., Elkassem, A. A., Park, J. M., Rothenberg, S. A., Tanwar, M., Yi, P. H., Sturdivant, A., Tridandapani, S., & Sotoudeh, H. (2023). Artificial intelligence literacy: Developing a multi-institutional infrastructure for AI education. *Academic Radiology, 30*(7), 1472–1480. https://doi.org/10.1016/j.acra.2022.10.002

PRISMA. (2020). PRISMA 2020 checklist. http://www.prisma-statement.org/documents/PRISMA_2020_checklist.pdf.

Purdue, U. Leading Ethically in the Age of AI and Big Data. Retrieved August 1, 2024, from https://cla.purdue.edu/about/college-initiatives/leadingethically/details.htmln.d.

Raji, I. D., Scheuerman, M. K., & Amironesei, R. (2021). You can't sit with us: Exclusionary pedagogy in AI ethics education. *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. https://doi.org/10.1145/3442188.3445914

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 conference on fairness, accountability, and transparency*. https://doi.org/10.1145/3351095.3372873

Reynolds, S. J. (2006). A neurocognitive model of the ethical decision-making process: Implications for study and practice. *Journal of Applied Psychology, 91*(4), 737–748. https://doi.org/10.1037/0021-9010.91.4.737

Sawyer, K. (2014). *The cambridge handbook of the learning sciences* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526

Sawyer, R. K. (Ed.). (2022). *The Cambridge handbook of the learning sciences* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781108888295.

Schiff, D. (2022). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education, 32*(3), 527–563. https://doi.org/10.1007/s40593-021-00270-2

Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020). What's next for AI ethics, policy, and governance? A global overview. *Proceedings of the AAAI/ACM conference on AI, ethics, and society*. https://doi.org/10.1145/3375627.3375804

Schiff, D. S., Kelley, S., & Camacho Ibáñez, J. (2024). The emergence of artificial intelligence ethics auditing. *Big Data & Society, 11*(4). https://doi.org/10.1177/20539517241299732

Shih, P.-K., Lin, C.-H., Wu, L. Y., & Yu, C.-C. (2021). Learning ethics in AI—teaching non-engineering undergraduates through situated learning. *Sustainability, 13*(7), 3718. https://doi.org/10.3390/su13073718

Southworth, J., Migliaccio, K., Glover, J., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers & Education: Artificial Intelligence, 4*, Article 100127. https://doi.org/10.1016/j.caeai.2023.100127

Stoyanovich, J. (2022). Teaching responsible data science. *1st International Workshop on Data Systems Education, 4–9*. https://doi.org/10.1145/3531072.3535318

Streveler, R. A., & Smith, K. A. (2020). Opinion: Course design in the time of coronavirus: Put on your designer's CAP. *8*(4).

Streveler, R. A., Smith, K. A., & Pilotte, M. (2012). Aligning course content, assessment, and delivery: Creating a context for outcome-based education. In K. M. Yusof, N. A. Azli, A. M. Kosnin, S. K. S. Yusof, & Y. M. Yusof (Eds.), *Outcome-based science, technology, engineering, and mathematics education: Innovative practices* (pp. 1–26). IGI Global. https://doi.org/10.4018/978-1-4666-1809-1.ch001.

Taylor, G., & Deb, D. (2021). Teaching AI ethics in a flipped classroom. *J. Inf. Comput. Sci. in colleges, 36*(5), 67–76.

Tenório, K., & Romeike, R. (2023). AI Competencies for non-computer science students in undergraduate education: Towards a competency framework. *Proceedings of the 23rd koli calling international conference on computing education research*. https://doi.org/10.1145/3631802.3631829

The White House. (2023). Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. *The White House*. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

Touretzky, D., Gardner-McCune, C., & Seehorn, D. (2022). Machine learning and the five Big Ideas in AI. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-022-00314-1

Tuovinen, L., & Rohunen, A. (2021). In *Teaching AI ethics to engineering students: Reflections on syllabus design and teaching methods*.

Umbrello, S., & Van De Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. *AI and Ethics, 1*(3), 283–296. https://doi.org/10.1007/s43681-021-00038-3

UNESCO. (2011). *International standard classification of education (ISCED) 2011*. UNESCO. https://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf.

UNESCO. (2022a). *International forum on AI and education: Synthesis report*. UNESCO.

UNESCO. (2022b). *Recommendations on the ethics of artificial intelligence*. UNESCO.

UNESCO. (2023). *ChatGPT and artificial intelligence in higher education*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000385146.

Vaismoradi, M., Jones, J., Turunen, H., & Snelgrove, S. (2016). Theme development in qualitative content analysis and thematic analysis. *Journal of Nursing Education and Practice, 6*(5), Article p100. https://doi.org/10.5430/jnep.v6n5p100

Van Brummelen, J., Heng, T., & Tabunshchyk, V. (2021). Teaching tech to talk: K-12 conversational artificial intelligence literacy curriculum and development tools. *Proceedings of the AAAI Conference on Artificial Intelligence, 35*(17), 15655–15663. https://doi.org/10.1609/aaai.v35i17.17844

Van den Beemt, A., MacLeod, M., Van der Veen, J., Van de Ven, A., van Baalen, S., Klaassen, R., & Boon, M. (2020). Interdisciplinary engineering education: A review of vision, teaching, and support. *Journal of Engineering Education, 109*(3), 508–555. https://doi.org/10.1002/jee.20347

Wang, L., Liu, Z., Liu, A., & Tao, F. (2021). Artificial intelligence in product lifecycle management. *The International Journal of Advanced Manufacturing Technology, 114*(3), 771–796. https://doi.org/10.1007/s00170-021-06882-1

Wang, L., Malhotra, D., & Murnighan, J. K. (2011). Economics education and greed. *The Academy of Management Learning and Education, 10*(4), 643–660. https://doi.org/10.5465/amle.2009.0185

Wang, G., Zhao, J., Van Kleek, M., & Shadbolt, N. (2024). Challenges and opportunities in translating ethical AI principles into practice for children. *Nature Machine Intelligence, 6*(3), 265–270. https://doi.org/10.1038/s42256-024-00805-x

Wiese, L., & Magana, A. (2024). A department's syllabi review for LLM considerations prior to university-standard guidance. *2024 ASEE annual conference & exposition proceedings. 2024 ASEE annual conference & exposition, portland, OR*. https://doi.org/10.18260/1-2–46436

Wiggins, G., & McTighe, J. (2005). What is backward design?. In *Understanding by design* (2nd ed., pp. 7–19). Assn. for Supervision & Curriculum Development.

Williams, R., Ali, S., Devasia, N., DiPaola, D., Hong, J., Kaputsos, S. P., Jordan, B., & Breazeal, C. (2022). AI + ethics curricula for middle school youth: Lessons learned from three project-based curricula. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-022-00298-y

Xia, B., Lu, Q., Zhu, L., Lee, S. U., Liu, Y., & Xing, Z. (2024). Towards a responsible AI metrics catalogue: A collection of metrics for AI accountability. *Proceedings of the IEEE/ACM 3rd international conference on AI engineering - software engineering for AI* (pp. 100–111). https://doi.org/10.1145/3644815.3644959

Young, J. J., & Annisette, M. (2009). Cultivating imagination: Ethics, education and literature. *Critical Perspectives on Accounting, 20*(1), 93–109. https://doi.org/10.1016/j.cpa.2007.03.003

Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2023). Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study. *International Journal of Artificial Intelligence in Education, 33*(2), 290–324. https://doi.org/10.1007/s40593-022-00293-3

Zhao, L., Wu, X., & Luo, H. (2022). Developing AI literacy for primary and middle school teachers in China: Based on a structural equation modeling analysis. *Sustainability, 14*(21), Article 14549. https://doi.org/10.3390/su142114549

Zhong, C.-B. (2011). The ethical dangers of deliberative decision making. *Administrative Science Quarterly, 56*(1), 1–25. https://doi.org/10.2189/asqu.2011.56.1.001

Zhou, Y., Zhan, Z., Liu, L., Wan, J., Liu, S., & Zou, X. (2022). International prospects and trends of artificial intelligence education: A content analysis of top-level AI curriculum across countries. *Proceedings of the 6th international conference on digital technology in education*. https://doi.org/10.1145/3568739.3568796