Conveying the Ethics of Artificial Intelligence in K–12 and Academia: A Systematic Review of Teaching Methods

Nils J. Tschoppe
Leuphana University of Lüneburg
nils.tschoppe@leuphana.de

Johannes W. Katsarov
Leuphana University of Lüneburg
johannes.katsarov@leuphana.de

Paul Drews
Leuphana University of Lüneburg
paul.drews@leuphana.de

Hannah Trittin-Ulbrich
Leuphana University of Lüneburg
hannah.trittin-ulbrich@leuphana.de

Abstract

Artificial intelligence (AI) and its recent advancements pervade vast areas of education, the workplace, and society. As a driver of technological progress, AI has the potential to transform entire business areas, optimize the way we work and live together, and promote creativity. Concomitantly, it harbors the risk of biased algorithms, discrimination, and misinformation. Accordingly, it is now more important than ever to teach students—as potential future designers and users of AI systems—how to deal responsibly and ethically with AI. To support educators in conveying the responsible and ethical use of AI, we conducted a systematic literature review based on the PRISMA guideline. As a result, we present an overview of established and innovative methods of teaching AI ethics in K–12 and academic settings. We discuss these in terms of their effectiveness and grounding in learning theories and derive implications for theory and practice.

Keywords: PRISMA, Artificial Intelligence, Ethics, Innovation, Education.

1. Introduction

The latest developments in the nascent field of artificial intelligence (AI)—such as generative AI—have a significant impact on what we teach (e.g., a curriculum design that aims at conveying both the fundamentals of AI and ethical concerns) and how we teach it (e.g., teaching methods such as case studies or dance). Educators, both from K–12 and academia, are entering uncharted territory as they face the challenge of shaping the teaching environment with effective and innovative methods that foster dealing responsibly and ethically with AI. Students—as potential future designers and users of AI—need to be taught in a responsible and ethical manner, emphasizing "human responsibility for the development of intelligent systems along fundamental human principles and values" (Dignum, 2019, p. 119).

In light of this increasing educational and societal relevance, Bleher & Braun (2023), Morley et al. (2020), and Schiff et al. (2020) stressed the need to intertwine ethical theory and practice. This has also recently been recognized by policymakers, illustrated by the multitude of public guidelines for ensuring the inclusion of ethical aspects of AI in education (European Commission, 2022; Holmes et al., 2022; Miao et al., 2021; Office of Educational Technology, 2023; Pons, 2023). In this context, Saltz et al. (2019) analyzed 186 courses across 20 programs teaching machine learning (ML) "as a subset of AI [and a] key component of everyday applications that mediate our social, cultural and political interactions" (Zammit et al., 2021, p. 1). In their systematic syllabi review, only 22 (12%) of the technical ML courses included ethics-related content stressing the need to transfer ethical principles to educational practice.

To illustrate the relevance of AI ethics for education and society, trolley scenarios can be used. They are "hypothetical but useful abstractions to consider how we want AI systems to behave in certain situations" (Dignum, 2019, p. 42). The "trolley problem" describes the moral dilemma of "the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed" (Foot, 1967, p. 8). It serves as an example to test "moral intuitions regarding the doctrine of double effect, Kantian principles and utilitarianism" (Andrade, 2019, p. 1) and can be transferred in a constructive manner to today's ethical challenges (e.g., addressing policymakers or developers on algorithmic decisions regarding autonomous vehicles).

Numerous examples of ethical challenges in the development and use of AI can also be found in other industries (e.g., law (Wright, 2020) and medicine (Gundersen & Bærøe, 2022)) and application areas (e.g., fairness and transparency in the recruiting process (Hunkenschroer & Luetge, 2022; van den Broek et al., 2019, 2021) or accountability in chemical research and

URI: https://hdl.handle.net/10125/109418
978-0-9981331-8-8

development (Hermann et al., 2021)), thereby stressing the importance of an interdisciplinary perspective on ethical AI in today's society.

With our study, we advance the fields of information systems, education, and ethics theoretically and practically through an interdisciplinary lens focused on supporting educators in conveying the ethical and responsible use of AI in K–12 and academic settings. Paying special attention to the plethora of teaching methods, the following research question guides our study:

Which established and innovative teaching methods support conveying the ethics of AI in K–12 and academia?

While we provide an overview of established and innovative teaching methods, we investigate which ethical risks and concerns were addressed and examine their effectiveness and the extent to which teaching methods were grounded in learning theory.

2. Theoretical Background

To provide educators with a better understanding of the status quo and latest trends in the development of AI ethics curricula, we first give an overview of selected in-depth systematic reviews. Second, we reflect on learning theory for structuring sessions and courses and highlight the need to evaluate the effectiveness of teaching methods.

In their unsupervised topic modeling analysis of global AI curricula on the teaching of tech ethics (including a crowd-sourced list of 166 academic courses and supplemental information on syllabi at undergraduate and graduate levels), Javed et al. (2022) revealed that ethics is predominantly taught in silos (e.g., specific topics or disciplines). They suggest developing hybrid ethics courses that combine interdisciplinary ethics modules (e.g., covering philosophy, accountability, or inequality) with domain-specific modules (e.g., taught by AI designers, engineers, or lawyers). To deepen this understanding, Fiesler et al. (2020) stressed the lack of standards and the high level of variability across the content of tech ethics courses. Based on an in-depth qualitative analysis of 115 syllabi from university technology ethics courses, they highlighted critical topics (e.g., privacy, algorithms, inequality/justice) within a set of 15 high-level categories. In addition, by systematically analyzing 31 standalone AI ethics classes and 20 AI/ML technical courses (which incorporated ethics) from the US, Garrett et al. (2020) provided a better understanding of the role of ethics in AI education, indicating that privacy, fairness, and bias were the most common elements when teaching ethics in technical courses.

Building on this, Hagendorff (2020), Huang et al. (2023), and Jobin et al. (2019) systematically screened AI ethics guidelines and underlying principles. In this process, Jobin et al. (2019, p. 389) indicated "a global convergence emerging around five ethical principles (transparency, justice and fairness, non-maleficence, responsibility and privacy)." More specific, while Adams et al. (2023) analyzed four AI ethics guidelines in their focus on K–12 (in terms of pedagogical appropriateness, children's rights, AI literacy, and teacher well-being), Memarian and Doleck (2023) provided an overview of qualitative and quantitative studies on fairness, accountability, transparency, and ethics in higher education.

However, despite the rich body of knowledge on AI ethics curricula, guidelines, and principles for educational purposes, questions concerning the selection and effectiveness of established and innovative teaching methods and strategies arise (Donker et al., 2014; Lee et al., 2021). Here, learning can be understood as "the process whereby knowledge is created through the transformation of experience" (Kolb, 1984, p. 38). This includes debates on the relevance of instructional guidance, contrasting popular minimal guided approaches (e.g., constructivist, inquiry-based, or experiential learning) with direct instructional guidance (Kirschner et al., 2006). While Kirschner et al. (2006) provided evidence for the superiority of the latter based on the human cognitive architecture and empirical studies, Kolb (2015) emphasized the relevance of experience as the source of learning in his experiential learning theory (ELT), which can support structuring sessions or courses using a learning cycle. In this context, Katsarov et al. (2022, p. 935) highlighted in their meta-analysis, that "experiential training approaches are more effective than classical case discussions." Moreover, they concluded that a practical emphasis on ethics education is often more effective than a primarily intellectual deliberation about ethical problems, regardless of the specific method.

Building on this, a better understanding of the effectiveness of established and innovative teaching methods contributes to transferring the ethics of AI from theory to practice.

3. Methodology

Our aim was to synthesize knowledge on established and innovative teaching methods for conveying the ethics of AI in academic and K–12 settings, paying attention to grounding teaching formats and methods in learning theory and evaluating their effectiveness. For this purpose, the systematic review is based on the guideline for Preferred Reporting Items for

Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009; Page et al., 2021). While it has been "designed primarily for systematic reviews of studies that evaluate the effects of health interventions" (Page et al., 2021, p. 2), it is well-established in other disciplines and supports to rigorously document the search process (Figure 1).

As we expected to find major contributions in leading journals, we followed the recommendation of Webster & Watson (2002) and screened journals related to the research question as the first step. Due to the interdisciplinary character of this study, we started searching in leading journals from the disciplines of information systems, education, and ethics. In terms of the information systems discipline, we relied on eight leading journals (European Journal of Information Systems, Information Systems Journal, Information Systems Research, Journal of Information Technology, Journal of Management Information Systems, Journal of the Association for Information Systems, Journal of Strategic Information Systems, and Management Information Systems Quarterly) to search for peer-reviewed articles in this field. Furthermore, as conference proceedings are crucial for and in information systems research, we included leading conferences, such as the International Conference on Information Systems and the European Conference on Information Systems, in the search process.

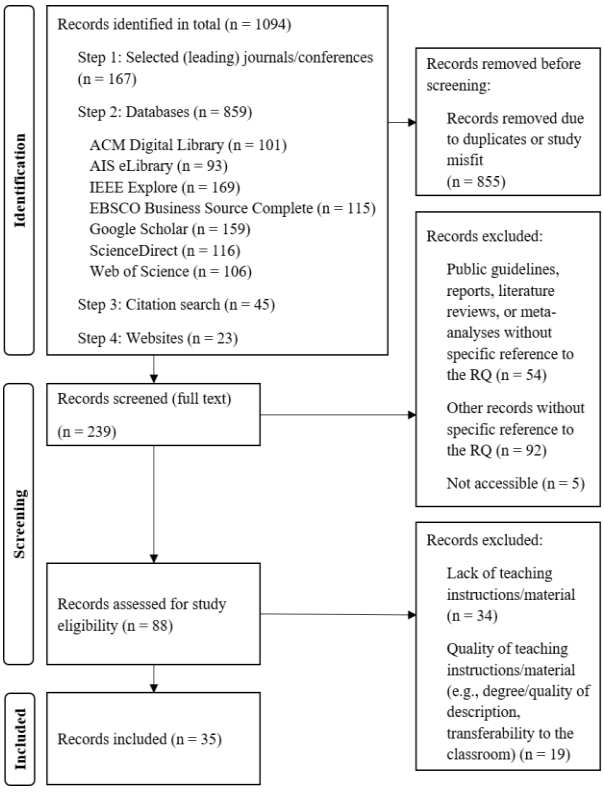

Figure 1. PRISMA flow chart

To include leading journals from the ethics and education discipline, we relied, among others, on the Clarivate Journal Citation Reports, which ranks journals according to impact and source metrics, such as the impact factor and total citations. Here, we included leading journals on ethics (Ethics and Information Technology, Ethics and Behavior, Journal of Responsible Innovation, Business Ethics Quarterly, Science and Engineering Ethics, and Journal of Business Ethics) and education (Academy of Management Learning and Education, Review of Educational Research, Educational Researcher, Review of Research in Education, Journal of Teacher Education, Computers and Education, and Computers and Education: Artificial Intelligence) that were likely to provide a solid base for our study.

The literature search was conducted November 07, 2023. We used the following "default" search string for identifying relevant literature focused on teaching formats and methods, subdivided into the three areas of AI, ethics, and education:

(AI OR "Artificial Intelligence" OR ML OR "Machine Learning") AND (Teach OR Litera* OR Educat* OR Learn* OR Syllab* OR Curricul* OR Train* OR Course OR School OR Universit* OR Academi*) AND (Ethic* OR Responsi* OR Trustworth* OR Human-Cent*).*

Concerning the selection of keywords in ethics, we followed Vassilakopoulou et al. (2022) and included responsible and trustworthy AI to grasp different streams of research on ethical AI. This included research on human-centered AI, which we included with an asterisk to reveal articles from the American and British literature. We also used asterisks for other keywords (e.g., to include teachers and teaching in our search using the search term teach*). There were no restrictions on coverage in terms of years.

Due to journal- and database-specific search functions (e.g., the length of the search string) and to reduce the number of unrelated articles, we narrowed our search to screen titles and abstracts in the first step and adjusted the search string accordingly. However, the "default" search string had to be adapted due to the individual search masks of the journals (e.g., searching for articles in Computers and Education through the ScienceDirect database did not support wildcards (*), or there were restrictions on the length, such as eight boolean connectors (per field). For example, as "there is consensus that machine learning is a subset of AI and that deep learning is a subset of machine learning" (Webb et al., 2021, p. 2112), we focused on the "core concept" AI in the search string and omitted the keywords ML and machine learning when screening the journals Computers and Education and Computers and Education: Artificial Intelligence, and the ScienceDirect

database in the subsequent search process. Since this adaptation of the search string harbored the risk of bias, we extended our research with scholarly databases to enable a comprehensive understanding of the curriculum design and teaching methods of ethical AI in education.

In the second step, we extended our literature review with research articles from scholarly databases, since most articles were found in the Science and Engineering Ethics journal and some others showed no results when we applying the default search string. Due to the novelty and increasing relevance of this topic for education, we also included non-peer-reviewed (grey) literature in this process. In doing so, we included the following databases: Association for Computing Machinery (ACM) Digital Library, the AIS eLibrary, IEEE Explore, ScienceDirect, EBSCO Business Source Complete, Web of Science, and (particularly for grey literature) Google Scholar. Due to a potentially high number of unrelated articles (searching for "ethics + AI + education" in the text field returned around three million results on Google Scholar) or restrictions on the length (ScienceDirect), we limited the search in selected databases. While we screened titles and abstracts by applying the (full) default search string in the AIS eLibrary, we focused on core concepts (e.g., by omitting the keywords ML and machine learning) in the Web of Science and Google Scholar databases and only searched in the titles. Although these variations in the search string harbored the risk of bias and not being exhaustive, they nevertheless enable a comprehensive understanding of established and innovative teaching methods for conveying the ethics of AI in K–12 and academia.

We then screened the identified records by scanning the titles and abstracts and applying initial eligibility criteria (e.g., regarding the study fit). Next, we performed a heuristic forward (identifying articles citing the key articles in the previous step) and backward (reviewing the citations for the articles identified in the first step) search on the most promising articles to identify related relevant articles, deemed to support a deeper understanding of teaching methods and supplementary material for education (e.g., online platforms, applications).

After removing duplicates, this resulted in 239 articles for which we screened the full texts.[1] In this process, we excluded other articles (e.g., records which were not accessible or without specific reference to the research question). This resulted in a list of 88 articles (including curricula, online platforms, and workshops) supporting a better understanding of the teaching methods used for conveying the ethics of AI. Finally, we included 35 articles from this list which seemed most promising with regard to the research question and were likely to be transferred to the classroom due to detailed teaching instructions and/or resources.

4. Findings

In the following, we provide an overview of established and innovative teaching methods that can support educators in conveying the ethics of AI in K–12 and academia (for a list of the records included, see Table 1 in the appendix). These aim at both a general audience (e.g., high school students and undergraduates with(-out) a technical background) as well as specific focus groups such as environmental scientists (AI2ES Center, 2023; McGovern et al., 2023) or black, indigenous, and other people of color (BIPoC) students (Lee et al., 2021; Payne, W. et al., 2021; Zhang et al., 2023). We problematize these findings paying attention to grounding these in learning theory and evaluating their effectiveness.

Teaching methods. First, the systematic literature review revealed a variety of didactic approaches on conveying the ethics of AI. Teaching through case studies was a popular method for teaching AI ethics through real-world events (e.g., Taylor & Deb, 2021) or science fiction (e.g., Burton et al., 2015, 2016, 2017), followed by discussions and debates on ethical AI, (e.g., Peace, 2011). Moreover, game-based approaches or gamification, both online and offline, were an up-and-coming topic to teach the fundamentals and ethics of AI. This was particularly evident in K–12, while the teaching of AI ethics through educational games in academia was in its infancy. For example, the ethics of AI curriculum for middle school students (Payne, B., 2019) included a comprehensive set of innovative teaching activities (e.g., AI Bingo, ethical matrix, Youtube scavenger hunt, supervised ML quiz), detailed instructions for teachers, and assessments. In addition, particularly in K–12, teaching children to train ML models was often contextualized in specific settings, such as the environment, and accompanied by teaching ethical awareness such as bias in algorithms (e.g., Code.org, 2020).

Second, in the previous process of screening the records for eligibility, a variety of other creative methods emerged which may serve as inspiration for designing courses, though these lacked detailed teaching instructions or resources, or aimed at teaching the ethics of AI outside the classroom (e.g., in a business context). For example, this included methods such as teaching AI ethics through dance (Castro et al., 2022; Walsh et al., 2022). Moreover, Tlili et al. (2023) used

[1] The list of the 239 records that were screened (full text) can be found here: *https://doi.org/10.5281/zenodo.13752065*

ChatGPT for generating learning content (e.g., quizzes or tests), while van Slyke et al. (2023) illustrated the potential of generative AI by prompting ChatGPT to write a case study. In addition, more specific approaches emerged, such as case studies on bias in face recognition (Orchard & Radke, 2023) or the use of the ECCOLA method and cards to support ethical thinking in the product development process from a business perspective (Vakkuri et al., 2021).

Third, when screening the records for eligibility, several articles lacked detailed or transferable teaching instructions but proved useful in terms of potentially enriching the teaching experience. For example, Stahl et al. (2023) provided 19 case studies for addressing ethical AI challenges from seven perspectives (unfair and illegal discrimination; privacy; surveillance capitalism; manipulation; right to life, liberty and security of persons; dignity; AI for good and the SDGs (sustainable development goals)). Moreover, the Association for Computing Machinery's Committee on Professional Ethics (2021) presents seven fictional case studies for practitioners viewed from the ACM Code of Ethics and Professional Practice. In addition, Adams et al. (2022) provided a typology and examples of AI education technologies for K–12, while Moorhouse et al. (2023) offered a list of generative AI tools used by leading universities. From a theoretical perspective, the Stanford Encyclopedia of Philosophy (Müller, 2020) and the bibliography of the Centre for Digital Philosophy (2013) may serve as a starting point for transferring the philosophical and ethical underpinnings of AI and robotics to the classroom. In the context of tools, Google's (2021) "Teachable Machine" (a web-based tool for creating ML models), or the "Moral machine" (a platform for gathering human perspective on moral decisions) (Max Planck Society for the Advancement of Science e.V., 2019) proved to be useful for educational purposes in K–12 and academia.

Learning theory. Our systematic literature review revealed that curricula and courses dealing with AI ethics were often insufficiently grounded in learning theory. However, we noticed several design-based or constructivist approaches to learning (e.g., Ali et al., 2019, 2023; Arastoopour Irgens et al., 2022; Henry et al., 2021; Zammit et al., 2021). For example, Zammit et al. (2021) introduced the educational game "ArtBot" to convey complex ML concepts (link can be found in the references). In doing so, they followed a constructivist approach to learning and built on exploratory learning principles to improve the AI literacy of K–12 students. Moreover, Henry et al. (2021) followed a design-based research (DBR) approach which was inspired by the game "Guess Who?." The curriculum aimed at conveying the basics of ML and AI ethics in general through a role-playing game, in which 10 to 14 years children alternate between the roles of an AI, a tester, or a developer. In contrast, Ng et al. (2022) used digital story writing as an inquiry-based pedagogical approach with five phases, which effectively fostered students' AI literacy. In the previous process of screening the records for eligibility, other insightful approaches to learning emerged. For example, Grøder et al. (2022) used Kolb's Experiential Learning Cycle to design a master's degree course about responsible AI use for public welfare services.

In the context of different approaches to learning, another question concerns what the teaching methods actually sensitized students to (e.g., which ethical theories, risks, and concerns were addressed). This covered major ethical theories such as virtue ethics (e.g., Bezuidenhout et al., 2020; Burton et al., 2015), deontological theories (e.g., Green, 2021), or the use of Kohlberg's Theory of Moral Development (Ali et al., 2019). Others were guided by ethical principles and related guidelines (e.g., Fong et al., 2022). This ranged from addressing social, legal, and ethical issues in general to principle-based approaches (e.g., Taylor & Deb (2021) introduced the Montreal Declaration of Responsible AI Development as a framework for ethical analysis) and the consideration of specific topics such as deepfakes or bias (e.g., Ali, DiPaola, Lee, et al., 2021).

While our findings align with Garrett et al. (2020), who indicate that privacy, fairness and bias were common elements in AI ethics education (see, e.g., AI4ALL, 2023; Code.org, 2020; Taylor & Deb, 2021), conveying AI ethics through ethical guidelines and frameworks was rather uncommon. In K–12, conveying the fundamentals of AI/ML was often linked to teaching specific ethical concerns such as bias in datasets, whereas in academia, AI ethics was comparatively more often taught separately from the fundamentals of AI/ML in our sample. Moreover, critical thinking (e.g., spotting deepfakes, misinformation and other manipulations) was commonly taught to foster students social media literacy (e.g., AI4ALL, 2023; Ali, DiPaola, & Breazeal, 2021; Ali, DiPaola, Lee, et al., 2021) or to sensitize for culturally situated issues such as racial or gender bias (e.g., Baumer et al., 2022; Taylor & Deb, 2021).

Effectiveness. Although teaching methods and curricula were evaluated in most of the studies (with the exception of online platforms), the individual evaluation parameters, approaches, and depth varied significantly.

For example, McGovern et al. (2023) reported positive results in their lecture in terms of students' understanding of the meaning of trustworthy AI, considering end users, or choosing suitable case studies in an environmental context. Moreover, in line with discussions and debates on ethical AI, which have been a successful teaching method in many (non-technical) disciplines (Peace, 2011), Burton et al. (2015) assessed

case studies in the field of science fiction positive outcomes in engaging students with critical thinking and AI ethics. However, they criticized the generalizability of their findings due to the small sample size, which leaves room for further research. In addition, Williams et al. (2022) emphasized the lack of validated summative assessments for middle school students, although the results of their workshops indicate positive outcomes in terms of students' understanding of ethical issues (e.g., privacy, responsible design, fairness) and provide insights into effective teaching strategies.

However, while these examples provide initial insights into the effectiveness of selected teaching methods and course formats, these are not sufficient for drawing general conclusions on the effectiveness or superiority of specific methods and, thus, for conveying the responsible and ethical use of AI in K–12 and academia. This manifested in a lack of detailed evaluation information, the small sample size, and individual evaluation parameters hampering comparability, which is critically for developing studies or meta-analyses that deal with teaching methods' suitability for conveying the ethics of AI. In particular, it needs to be questioned which method is promising for teaching specific knowledge (e.g., conceptual understanding, critical thinking, ethical awareness) in order to promote the ethical AI literacy of students.

Thus, as the rapid technology advancement is creating new opportunities for developing innovative teaching formats (such as dancing with AI), paying greater attention to existing learning theories may guide and enrich these and foster students' learning success and motivation. This may also involve future efforts to achieve a higher degree of comparability in terms of standardized evaluation criteria. In this context, future research may focus on examining differences in the effectiveness of minimally guided approaches (e.g., constructivist learning, inquiry learning, experiential learning) and direct instructional guidance. In particular, by following Katsarov et al. (2022) and McNamara et al. (2018), the concept of teaching AI ethics through codes of ethics, guidelines, or a primarily intellectual deliberation about ethical problems needs to be questioned in terms of their effectiveness.

5. Discussion, Limitations, and Future Research

Due to recent advancements in AI and its far-reaching impact on society, it is crucial to prepare and sensitize students—as the next generation of technologists—in the responsible and ethical use and building of AI systems. With our study, we respond to this increasing relevance and provide guidance for educators in conveying the ethics of AI. The systematic literature search based on the PRISMA guideline provides an overview of the current state of research on established and innovative teaching methods.

Stressing the need to put theory of AI ethics into practice, we provide three future avenues for enriching research and teaching experience: 1) strengthening the grounding of teaching formats and methods in learning theory, 2) a greater focus on the effectiveness of teaching methods, and 3) increased openness to innovative approaches. Referring to the latter implication, we noticed that K–12 played a pioneering role in terms of the richness of innovative approaches to teach AI ethics. Accordingly, we plead for more academic openness and creative teaching formats and methods (e.g., through arts or the use of online and offline game-based approaches) in the context of teaching the ethics of AI as an opportunity for enhancing teaching and learning experience.

However, this study also has some limitations. While the purpose of the study was to support educators in conveying the ethics of AI by providing an overview of established and innovative teaching methods, the boundaries between curricula, which often contained several teaching methods, and research on specific teaching methods were blurred. Here, both foci allow for independent in-depth analysis. For example, a more comprehensive guide on curriculum design, divided into K–12 and academic settings, seems to be promising. This also relates to the need to further investigate teaching methods in general or, for example, game-based approaches in particular. Accordingly, the search process and related terms need to be adapted to also include articles from other disciplines that can be transferred to teaching the ethics of AI to students (e.g., by including "gamification" and related terms in the search string). Moreover, the distinction between eligible and included articles (e.g., with regard to the transferability for the development of own teaching formats) was to some extent blurry depending on the target audience. We also noticed that the Massachusetts Institute of Technology (MIT) played a pioneering role in K–12 research on ethical AI, which leaves room for investigating different approaches to the development of curricula and teaching methods at a global level.

Another critique is rooted in the use of grey literature, which is grounded in the novelty of the topic but also has implications for future research on AI ethics in education. We addressed this novelty by combining curricula and teaching methods for K–12 and academia. While separating these domains would support a clear distinction, it also illustrates the interdisciplinary nature and multi-faceted application of methods and tools that can be meaningfully transferred from one educational area to another.

6. References

Adams, C., Pente, P., Lemermeyer, G., & Rockwell, G. (2023). Ethical Principles for Artificial Intelligence in K-12 Education. *Computers and Education: Artificial Intelligence*, *4*(1), 1–10.

Adams, C., Pente, P., Lemermeyer, G., Turville, J., & Rockwell, G. (2022). Artificial Intelligence and Teachers' New Ethical Obligations. *The International Review of Information Ethics*, *31*(1), 1–18.

AI2ES Center. (2023). *AI2ES – Educational Resources for AI and Machine Learning*. https://www.ai2es.org/products/education/

AI4ALL. (2023). *AI4ALL Open Learning*. https://ai-4-all.org/resources/

Ali, S., DiPaola, D., & Breazeal, C. (2021). What are GANs?: Introducing Generative Adversarial Networks to Middle School Students. *35th AAAI Conference on Artificial Intelligence*.

Ali, S., DiPaola, D., Lee, I., Sindato, V., Kim, G., Blumofe, R., & Breazeal, C. (2021). Children as Creators, Thinkers and Citizens in an AI-driven Future. *Computers and Education: Artificial Intelligence*, *2*(1), 1–11.

Ali, S., DiPaola, D., Williams, R., Ravi, P., & Breazeal, C. (2023). Constructing Dreams Using Generative AI. *Preprint*, 1–15.

Ali, S., Payne, B., Williams, R., Won Park, H., & Breazeal, C. (2019). Constructionism, Ethics, and Creativity: Developing Primary and Middle School Artificial Intelligence Education. *28th International Joint Conference on Artificial Intelligence*.

Andrade, G. (2019). Medical Ethics and the Trolley Problem. *Journal of Medical Ethics and History of Medicine*, *12*(3), 1–15.

Arastoopour Irgens, G., Vega, H., Adisa, I., & Bailey, C. (2022). Characterizing Children's Conceptual Knowledge and Computational Practices in a Critical Machine Learning Educational Program. *International Journal of Child-Computer Interaction*, *34*(1), 1–15.

Association for Computing Machinery's Committee on Professional Ethics. (2021). *ACM Ethics – Case Studies*. https://ethics.acm.org/integrity-project/case-studies/

Bates, R. A. (2011). AI & SciFi: Teaching Writing, History, Technology, Literature and Ethics. *118th Annual American Society for Engineering Education Conference and Exposition*.

Baumer, B. S., Garcia, R. L., Kim, A. Y., Kinnaird, K. M., & Ott, M. Q. (2022). Integrating Data Science Ethics Into an Undergraduate Major: A Case Study. *Journal of Statistics and Data Science Education*, *30*(1), 15–28. https://smithcollege-sds.github.io/sds-www/ethics.html

Bezuidenhout, L., Quick, R., & Shanahan, H. (2020). "Ethics When You Least Expect It": A Modular Approach to Short Course Data Ethics Instruction. *Science and Engineering Ethics*, *26*(4), 2189–2213.

Bleher, H., & Braun, M. (2023). Reflections on Putting AI Ethics into Practice: How Three AI Ethics Approaches Conceptualize Theory and Practice. *Science and Engineering Ethics*, *29*(3), 1–21.

Burton, E., Goldsmith, J., Koenig, S., Kuipers, B., Mattei, N., & Walsh, T. (2017). Ethical Considerations in Artificial Intelligence Courses. *AI Magazine*, *38*(2), 1–29.

Burton, E., Goldsmith, J., & Mattei, N. (2015). Teaching AI Ethics Using Science Fiction. *Workshops of the 29th AAAI Conference on Artificial Intelligence*.

Burton, E., Goldsmith, J., & Mattei, N. (2016). Using "The Machine Stops" for Teaching Ethics in Artificial Intelligence and Computer Science. *Workshops of the 30th AAAI Conference on Artificial Intelligence*.

Castro, F. E. V., DesPortes, K., Payne, W., Bergner, Y., & McDermott, K. (2022). AI + Dance: Co-Designing Culturally Sustaining Curricular Resources for AI and Ethics Education Through Artistic Computing. *18th ACM Conference on International Computing Education Research*.

Centre for Digital Philosophy. (2013). *PhilPapers*. https://philpapers.org/browse/ethics-of-artificial-intelligence

Code.org. (2020). *AI for Oceans*. https://code.org/oceans

CS Frontiers. (2022). *AI and Machine Learning*. https://csfrontiers.org/ai-and-machine-learning.html

Dignum, V. (2019). *Responsible Artificial Intelligence*. Springer Nature.

Donker, A. S., de Boer, H., Kostons, D., Dignath van Ewijk, C. C., & van der Werf, M. P. C. (2014). Effectiveness of Learning Strategy Instruction on Academic Performance: A Meta-Analysis. *Educational Research Review*, *11*(1), 1–26.

European Commission. (2022). *Ethical Guidelines on the Use of Artificial Intelligence (AI) and Data in Teaching and Learning for Educators*. Publications Office of the European Union.

Fiesler, C., Garrett, N., & Beard, N. (2020). What Do We Teach When We Teach Tech Ethics? *51st ACM Technical Symposium on Computer Science Education*.

Fong, A., Gupta, A., Carr, S., Bhattacharjee, S., & Harnar, M. (2022). Modular Experiential Learning for Secure, Safe, and Reliable AI: Curricular Initiative to Promote Education in Trustworthy AI. *23rd Annual Conference on Information Technology Education*. https://wmich.edu/cs/cybertraining

Foot, P. (1967). The Problem of Abortion and the Doctrine of the Double Effect. *The Oxford Review*, *5*(1), 5–15.

Garrett, N., Beard, N., & Fiesler, C. (2020). More Than "If Time Allows": The Role of Ethics in AI Education. *3rd AAAI/ACM Conference on AI, Ethics, and Society*.

Google LLC. (2021). *Teachable Machine*. https://teachablemachine.withgoogle.com/

Green, N. (2021). An AI Ethics Course Highlighting Explicit Ethical Agents. *4th AAAI/ACM Conference on AI, Ethics, and Society*.

Grøder, C. H., Parmiggiani, E., Pappas, I., Schmager, S., Vassilakopoulou, P., & Papavlasopoulou, S. (2022). Educating about Responsible AI in IS: Designing a Course Based on Experiential Learning. *43rd International Conference on Information Systems*.

Gundersen, T., & Bærøe, K. (2022). The Future Ethics of

Artificial Intelligence in Medicine: Making Sense of Collaborative Models. *Science and Engineering Ethics*, *28*(2), 1–16.
Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, *30*(1), 99–120.
Hardebolle, C., Kovacs, H., Simkova, E., Pinazza, A., Di Vincenzo, M. C., Jermann, P., Tormey, R., & Dehler Zufferey, J. (2022). A Game-Based Approach To Develop Engineering Students' Awareness About Artificial Intelligence Ethical Challenges. *50th Annual Conference of the European Society for Engineering Education*. https://go.epfl.ch/MLethicsgame
Henry, J., Hernalesteen, A., & Collard, A. S. (2021). Teaching Artificial Intelligence to K-12 Through a Role-Playing Game Questioning the Intelligence Concept. *KI - Künstliche Intelligenz*, *35*(2), 171–179.
Hermann, E., Hermann, G., & Tremblay, J. C. (2021). Ethical Artificial Intelligence in Chemical Research and Development: A Dual Advantage for Sustainability. *Science and Engineering Ethics*, *27*(4), 1–16.
Holmes, W., Persson, J., Chounta, I.-A., Wasson, B., & Dimitrova, V. (2022). *Artificial Intelligence and Education – A Critical View through the Lens of Human Rights, Democracy and the Rule of Law*. Council of Europe Publishing.
Huang, C., Zhang, Z., Mao, B., & Yao, X. (2023). An Overview of Artificial Intelligence Ethics. *IEEE Transactions on Artificial Intelligence*, *4*(4), 799–819.
Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda. *Journal of Business Ethics*, *178*(4), 977–1007.
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*(1), 933–965.
Jobin, A., Ienca, M., & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.
Katsarov, J. W., Andorno, R., Krom, A., & van den Hoven, M. (2022). Effective Strategies for Research Integrity Training—a Meta-analysis. *Educational Psychology Review*, *34*(2), 935–955.
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75–86.
Kolb, D. A. (1984). *Experiential Learning: Experience as The Source of Learning and Development*. Prentice Hall.
Kolb, D. A. (2015). *Experiential Learning: Experience as The Source of Learning and Development*. Pearson Education.
Lane, D. (2018). *Machine Learning for Kids*. https://machinelearningforkids.co.uk/
Lee, I., Ali, S., Zhang, H., Dipaola, D., & Breazeal, C. (2021). Developing Middle School Students' AI Literacy. *52nd ACM Technical Symposium on Computer Science Education*.
Max Planck Society for the Advancement of Science e.V. (2019). *Moral Machine*. https://www.moralmachine.net/hl/de
McGovern, A., John, D., Ii, G., Wirz, C. D., Ebert-Uphoff, I., Bostrom, A., Rao, Y., Schumacher, A., Flora, M., Chase, R., Mamalakis, A., Mcgraw, M., Lagerquist, R., Redmon, R. J., & Peterson, T. (2023). Trustworthy Artificial Intelligence for Environmental Sciences: An Innovative Approach for Summer School. *Bulletin of the American Meteorological Society*, *104*(6), 1222–1231.
McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's Code of Ethics Change Ethical Decision Making in Software Development? *26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 729–733.
Memarian, B., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and Higher Education: A Systematic Review. *Computers and Education: Artificial Intelligence*, *5*(1), 1–12.
Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and Education: Guidance for Policy-makers*. UNESCO.
MIT Media Lab. (2023). *Dancing with AI*. https://dancingwithai.media.mit.edu/curriculum
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J. J., Devereaux, P. J., Dickersin, K., Egger, M., Ernst, E., Gøtzsche, P. C., … Tugwell, P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, *6*(7), 1–6.
Moorhouse, B. L., Yeo, M., Luke, B., Alina, M., & Wan, Y. (2023). Generative AI Tools and Assessment: Guidelines of the World' s Top-Ranking Universities. *Computers and Education Open*, *5*(1), 1–11.
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.
Müller, V. C. (2020). *Ethics of Artificial Intelligence*. https://plato.stanford.edu/entries/ethics-ai/
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 and Education: Artificial Intelligence*, *3*(1), 1–14.
Office of Educational Technology. (2023). *Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations*. U.S. Department of Education.
Orchard, A., & Radke, D. (2023). An Analysis of Engineering Students' Responses to an AI Ethics

Scenario. *37th AAAI Conference on Artificial Intelligence*.

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. *BMJ*, *372*(71), 1–9.

Payne, B. (2019). *An Ethics of Artificial Intelligence Curriculum for Middle School Students*. https://ec.europa.eu/futurium/en/system/files/ged/mit_ai_ethics_education_curriculum.pdf

Payne, W., Bergner, Y., West, M. E., Charp, C., Shapiro, R. B., Szafr, D. A., Taylor, E. V., & DesPortes, K. (2021). danceOn: Culturally Responsive Creative Computing. *CHI Conference on Human Factors in Computing Systems*.

Peace, A. G. (2011). Using Debates to Teach Information Ethics. *Journal of Information Systems Education*, *22*(3), 233–237.

Pons, A. (2023). *Generative AI in the Classroom: From Hype to Reality?* OECD Publishing.

Rueben, M. (2022). *Exploring the Future Through Robotics and Artificial Intelligence*. https://osf.io/tua5k

Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., Dewar, N., & Beard, N. (2019). Integrating Ethics within Machine-learning Courses. *ACM Transactions on Computing Education*, *19*(4), 1–26.

Schiff, D., Rakova, B., Ayesh, A., Fanti, A., & Lennon, M. (2020). Principles to Practices for Responsible AI: Closing the Gap. *Workshops of the 24th European Conference on AI*.

Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). *Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges*. Springer Nature.

Taylor, G., & Deb, D. (2021). Teaching AI Ethics in a Flipped Classroom. *Journal of Computing Sciences in Colleges*, *36*(5), 67–76.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the Devil Is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education. *Smart Learning Environments*, *10*(1), 1–24.

Vakkuri, V., Kemell, K. K., Jantunen, M., Halme, E., & Abrahamsson, P. (2021). ECCOLA — A Method for Implementing Ethically Aligned AI Systems. *Journal of Systems and Software*, *182*(3), 1–16.

van den Broek, E., Sergeeva, A., & Huysman Vrije, M. (2019). Hiring Algorithms: An Ethnography of Fairness in Practice. *40th International Conference on Information Systems*.

van den Broek, E., Sergeeva, A., & Huysman Vrije, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS Quarterly*, *45*(3), 1557–1580.

van Slyke, C., Johnson, R. D., & Sarabadani, J. (2023). Generative Artificial Intelligence in Information Systems Education: Challenges, Consequences, and Responses. *Communications of the Association for Information Systems*, *53*(1), 1–21.

Vassilakopoulou, P., Parmiggiani, E., Shollo, A., & Grisot, M. (2022). Responsible AI: Concepts, Critical Perspectives and an Information Systems Research Agenda. *Scandinavian Journal of Information Systems*, *34*(2), 89–112.

Walsh, B., Ali, S., Castro, F., Desportes, K., DiPaola, D., Lee, I., Payne, W., Sieke, S., & Zhang, H. (2022). Making Art with and about Artificial Intelligence: Three Approaches to Teaching AI and AI Ethics to Middle and High School Students. *53rd Annual SIGCSE Technical Symposium on Computer Science Education*.

Webb, M. E., Fluck, A., Magenheim, J., Malyn-Smith, J., Waters, J., Deschênes, M., & Zagami, J. (2021). Machine Learning for Human Learners: Opportunities, Issues, Tensions and Threats. *Educational Technology Research and Development*, *69*(4), 2109–2130.

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), 13–23.

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*, *33*(2), 325–383.

Williams, R., Kaputsos, S. P., & Breazeal, C. (2021). Teacher Perspectives on How To Train Your Robot A Middle School AI and Ethics Curriculum. *35th AAAI Conference on Artificial Intelligence*.

Williams, R., Park, H. W., Oh, L., & Breazeal, C. (2019). Popbots: Designing an Artificial Intelligence Curriculum for Early Childhood Education. *9th AAAI Symposium on Educational Advances in Artificial Intelligence*.

Wright, S. A. (2020). AI in the Law: Towards Assessing Ethical Risks. *8th IEEE International Conference on Big Data*.

York, E., & Conley, S. N. (2020). Creative Anticipatory Ethical Reasoning with Scenario Analysis and Design Fiction. *Science and Engineering Ethics*, *26*(6), 2985–3016.

Zammit, M., Voulgari, I., Liapis, A., & Yannakakis, G. (2021). The Road to AI Literacy Education: From Pedagogical Needs to Tangible Game Design. *15th European Conference on Games Based Learning*. http://learnml.eu/artbot.php

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://raise.mit.edu/daily/

7. Appendix

Table 1. Teaching methods (records included)

#	Author(s)	Setting	Target audience	Domain	Teaching method[*1]								Topic[*2]					Resource[*3]				
					C	P	S	R	G	A	V	O	F	C	P	A	C	W	A	O	T
1-2	AI2ES Center (2023); McGovern et al. (2023)	Academia	Early-career scientists	Environmental sciences	x		x				x	x	x	x	x		x	x	x	x	x
3	AI4ALL (2023)	K–12	High school	Interdisciplinary		x				x	x	x	x	x	x	x	x		x	x	x
4	Ali et al. (2023)	K–12	High school	Arts and culture					x	x			x	x		x		x			x
5-8	Ali et al. (2019); Williams et al. (2019, 2021, 2022)	K–12	Middle school	Arts and culture		x	x	x	x	x		x	x	x		x	x	x	x	x	x
9-10	Ali, DiPaola, Lee, et al. (2021); Ali, DiPaola, & Breazeal (2021)	K–12	Middle school	Social media		x	x		x		x		x	x		x	x	x	x	x	
11	Arastoopour Irgens et al. (2022)	K–12	Age 9 to 13	Arts and culture		x	x	x		x	x	x	x	x		x			x	x	
12	Bates (2011)	Academia	Undergraduates, graduates	Science fiction	x						x			x				x	x		
13	Baumer et al. (2022)	Academia	Undergraduates (data science)	Data ethics	x		x					x		x	x	x		x	x	x	x
14	Bezuidenhout et al. (2020)	Academia	Not specified	Data ethics	x		x					x	x	x	x	x	x	x	x		
15-16	Burton et al. (2015, 2016)	Academia	Undergraduates (computer science)	Science fiction	x						x			x					x		
17	Burton et al. (2017)	Academia	Not specified	Science fiction	x						x			x	x	x					x
18	Payne, W. et al. (2021)	K–12	BIPoC students	Arts and culture, empowerment			x			x	x		x			x				x	x
19	Code.org (2020)	K–12	Grade 3 to high school	Environment		x	x				x	x	x	x		x	x	x	x	x	x
20	CS Frontiers (2022)	K–12	Computer science	Interdisciplinary			x			x	x	x	x	x		x	x	x	x	x	x
21	Fong et al. (2022)	Academia	Undergraduates, graduates (STEM)	Cyber-infrastructure			x						x	x	x			x	x	x	x
22	Green (2021)	Academia	Undergraduates (computer science)	Ethical agents	x		x	x			x	x		x	x	x		x			
23	Hardebolle et al. (2022)	Academia	Graduates (engineering)	Gamification	x	x								x				x		x	x
24	Henry et al. (2021)	K–12	Age 10 to 14	Gamification		x						x	x		x	x	x		x	x	
25	Lane (2018)	K–12	Not specified	ML models		x	x			x	x	x	x	x					x	x	x
26-27	Lee et al. (2021); Zhang et al. (2023)	K–12	Middle school (students of color)	Interdisciplinary, empowerment	x	x	x		x	x	x	x	x	x		x	x	x	x	x	x
28	MIT Media Lab (2023)	K–12	Not specified	Arts and culture		x	x			x	x	x	x	x		x	x	x	x	x	x
20	Ng et al. (2022)	K–12	Primary school	Arts and culture	x	x	x	x		x	x	x	x	x	x	x			x		
30	Payne, B. (2019)	K–12	Middle school (grades 5 to 8)	Interdisciplinary		x	x		x	x	x	x	x	x		x	x	x	x	x	x
31	Peace (2011)	Academia	Management Information Systems (MIS) major	Information ethics	x											x		x			
32	Rueben (2022)	Academia	Not specified	Interdisciplinary	x	x	x	x			x	x	x	x	x	x	x	x	x	x	x
33	Taylor & Deb (2021)	Academia	Business analytics	Interdisciplinary	x									x	x			x			
34	York & Conley (2020)	Academia	Undergraduates (STEM)	Design fiction	x							x		x	x				x		
35	Zammit et al. (2021)	K–12	Primary and secondary school	Gamification		x					x		x	x		x			x	x	

[*1] C = case studies, debates; P = games; S = software, coding; R = robotics, chatbots; G = generative AI; A = art (drawing, writing, dancing); V = video, movie; O = other

[*2] F = fundamentals/mechanisms of AI/ML; C = concerns (ethical, social, legal); P = ethical principles/frameworks; A = ethical awareness, critical thinking

[*3] C = curriculum; W = workshop, course, module; A = learning activities, examples; O = online platform, website/link; T = teaching slides, worksheets