Ethical framework for AI in education

Bhoomika Agarwal

Open Universiteit, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands

Abstract

As Artificial Intelligence (AI) becomes increasingly integrated into various facets of our lives, including the educational domain, it is important to apply ethical principles to guide the development and deployment of AI systems. This ethically guided approach aims to mitigate potential harms or discriminatory outcomes resulting from AI algorithms. As a result, various ethical regulations and guidelines for AI ethics have emerged at the corporate, national, and supranational levels. However, the literature has paid relatively scant attention to the specific ethical considerations within the domain of AI in Education (AIED). AIED ethics represents a complex intersection, necessitating the combination of general AI ethics and the ethics of educational technology. This research aims to find the key constituents of an ethical framework for educational stakeholders of AIED that can be used to identify ethical issues in an AIED system. In this paper, we outline the methodology employed in this research to create an ethical framework for AIED. A systematic literature review will first look into the ethics of AI, the ethics of education and the ethics of AIED to consolidate the Ethical Values (EVs) and Ethical norms (ENs) for AIED ethics. Building on this knowledge from the literature, additional ENs will be collected through stakeholder consultation. These ENs will then be ranked by experts and used to form an ethical framework.

Keywords

Artificial Intelligence, Education, Ethics

1. Introduction

Artificial Intelligence (AI) is reshaping the world in profound ways and has a widespread impact on our lives, including education. The usage of AI in classrooms and in education is promising and provides opportunities to improve the education process. AI has been applied in educational contexts in a range of contexts varying from automation of administrative processes and tasks to curriculum and content development, instruction to understanding and improving students' learning processes through analysis of student data [1].

Over the past decade, the use of AI tools to support and enhance learning has grown exponentially [2]. In a recent literature review, Chen et al. looked at 20 years of AIED from 2000 to 2019 and shared several relevant findings: (a) AIED has seen an increased interest due of the positive effect of AI on learning; (b) there is an increase in AIED literature over the years; (c) AIED research is especially found in interdisciplinary journals with a dual focus on education and technology [3]. With the increased interest in AIED, there is a need to ethically guide the usage of AIED systems. The EU AI Act classifies the usage of AIED as 'high-risk' as "such systems may violate the right to education and training as well as the right not to be discriminated against and perpetuate historical patterns of discrimination" [4, p. 26].

Proceedings of the Doctoral Consortium of the Nineteenth European Conference on Technology Enhanced Learning, September 16–20, 2024, Krems, Austria

*Corresponding author.

bhoomika.agarwal@ou.nl (B. Agarwal)

0000-0002-7347-8465 (B. Agarwal)

An ethical guidance for AIED is necessary to reduce the negative impacts caused due to propagation of historical biases and discrimination that can result from the usage of AIED. At the same time, it is important to protect the privacy and autonomy of students and teachers so that the data collected by educational institutes cannot be used for other purposes. Hence, there is a need to use an ethical framework to regulate AIED.

In addition to considering general AI ethics, AIED ethics has to also consider the ethics of education. The overlap between the ethics of AI, ethics of education and ethics of AIED suggests that they should draw inspiration from each other [5]. The usage of AI technologies in education raises questions linked to ethical issues such as data ownership and control, privacy, biases in algorithms, data management, transparency, and a need for educational context [5]. Despite these concerns raised by AIED systems, limited attention has been paid to the ethics of AIED [6, 5, 7, 8, 9, 10]. AIED ethics could borrow from both domains and add additional ethical values as required by the domain specifically, while also considering the applicability of these values to the domain of AIED [5]. Due to these complexities, AIED ethics deserves attention.

The main research question guiding this research is: "What are the key constituents of an ethical framework for educational stakeholders of AIED that should be used to identify ethical issues in an AIED system?". This can be divided into three sub-questions for the three groups of educational stakeholders in AIED - students, educators and educational institutes. Thus, the three sub research questions are:

1. RQ1: What are the key constituents of an ethical

framework for educational institutes of AIED that should be used to identify ethical issues in an AIED system?
2. RQ2: What are the key constituents of an ethical framework for educators of AIED that should be used to identify ethical issues in an AIED system?
3. RQ3: What are the key constituents of an ethical framework for students of AIED that should be used to identify ethical issues in an AIED system?

The next section explains the methodology followed to answer these research questions.

2. Methodology

In order to identify the key constituents of an ethical framework for AIED, this research follows an adapted version of the generalizable model for developing Codes of Practice (CoP) developed by Sclater. The 'CoP for Learning Analytics' developed by the authors contains a set of guidelines and some ethical principles [11]. Due to the similarity of the format of the developed CoP to an ethical framework and the systematic nature of the generalizable methodology, we chose to use this approach to develop our ethical framework for AIED.

The author presents the activities followed as a basis for a generalizable model that can be used for developing CoP in other professions or areas of education [11]. The generalizable model involves developing five products: 1) a literature review identifying ethical, legal, and logistical concerns, 2) a taxonomy of issues refined through expert consultation, 3) a draft CoP, 4) a final publicly released CoP incorporating feedback from public consultation, and 5) a supporting website with guidance and case studies [11]. An advisory group of experts and stakeholders provides input throughout. As the CoP is piloted, feedback informs updates to subsequent versions.

This research uses the generalizable model developed by Sclater but makes three modifications. Firstly, we start with consulting stakeholders and incorporating their input from the beginning. This is because the ethics of AIED is a relatively new domain and has a direct impact on the lives of the stakeholders. Secondly, the experts are involved in a later stage of the model to review the input collected from the stakeholders. Additionaly, we gather input about the design of the framework from the experts. Lastly, due to time limitations, we do not follow the last two steps described in the generalizable model. Figure 1 shows adapted the methodology followed in this research.

To answer the main research question, a systematic literature review (SLR) that looks into the ethics of AI, the ethics of education and the ethics of AIED will first be conducted. This SLR will serve as the foundation by identifying the key constituents for an ethical framework for AIED. Following this, the gaps in literature will be filled in through stakeholder consultation. The output of the stakeholder consultation will then be ranked by experts and used to form an ethical framework. The stakeholder consultation and expert consultation will ensure that the viewpoints of all the stakeholders of AIED can be incorporated into the framework.

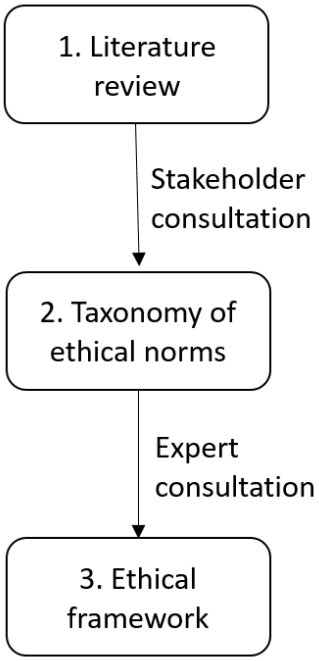

Figure 1: Research methodology

2.1. Systematic literature review

The existing literature on AIED ethics is lacking in three aspects: (a) a theoretical definition of the essence of AIED ethics, (b) a hierarchical classification of AIED ethics, and (c) reflection on the regulations [12]. As the first step of this research, a systematic literature review (SLR) was conducted with an aim to address these gaps by identifying and defining the key constituents of AIED ethics.

In this SLR, the normative approach to ethics - which proposes "how to act, how to live and what kind of person to be" [13, p. 2] - was followed. Ethics is defined in terms of values and norms. Values are abstract ideas that are strived for via certain types of behavior while norms are rules that specify actions to achieve certain values [14]. Using this definition of ethics, the key constituents of an ethical framework are identified as ethical values (EVs) and ethical norms (ENs). The SLR was guided by two

central research questions:

1. What are the main ethical values (EVs) for AIED?
2. What are the ethical norms (ENs) for AIED?

This SLR aims to identify the main EVs and ENs for AIED by analyzing relevant literature published between 2010 and 2022 that is available in the English language. Through a comprehensive database search and backward snowballing technique, a total of 25 articles were identified and included for analysis. The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) method [15] was used to ensure transparency and reproducibility of the SLR.

To identify the EVs, the definitions found within the literature were collected and reported. To analyse the EVs, they were grouped into common themes or topics. The grouping criterion was based on the identification of common key terms that appear consistently across multiple definitions of the EVs. Subsequently, the emergent themes were assigned descriptive labels that encapsulated the overarching meaning and conceptual essence conveyed by the set of definitions within each respective cluster. This thematic analysis process allowed for a systematic consolidation and organization of the EVs identified in the literature. The analysis of the EVs revealed six main EVs of AIED: non-discrimination, data stewardship, human oversight, goodwill, explicability, and educational aptness.

ENs were identified by looking for keywords such as 'norms', 'guidelines', 'regulations', 'recommendations' or another synonym of these terms in the selected articles. Four main stakeholder groups were identified from the literature: end users, developers, regulators and educational institutes [16]. The ENs identified in the literature were categorized according to the stakeholder groups they are relevant for and the corresponding main EV they uphold. Subsequently, these two categorizations were combined into a matrix, mapping ENs for specific stakeholder groups to the implementation of particular main EVs. This result could be used to provide the ENs for specific stakeholder sets to implement a given main EV.

In addition to answering the two research questions, the following points of discussion were highlighted by the SLR:

1. Ethics should be included in the design of AIED
2. Ethics of AIED should focus on the educational aptness of AIED solutions
3. More ENs should be established for end users of AIED
4. There exists a tight coupling between EVs, leading to possible ethical dilemmas

The main EVs and ENs identified in this SLR could serve as a foundation for developing an ethical framework for AIED. This SLR distinguished the key constituents that would comprise such an ethical framework, thereby setting the stage for further elaboration and fleshing out of these key constituents.

2.2. Stakeholder consultation

This research study aims to explore the ethical perspectives of key stakeholders in the field of AIED. The SLR highlighted a lack of ENs for end users and educational institutes. Building on this result, the target population for the stakeholder consultation comprised three distinct roups: students, teachers, and educational administrators who act as representatives from educational institutions.

To facilitate an in-depth exploration of AIED ethics, the study employs a qualitative research methodology involving focus group discussions. For this study, participants were recruited based on the match of their profile with the stakeholder groups identified, i.e. educational administrators, students and educators active in higher education. The participants were approached during workshops, presentations at conferences, summer schools, and other academic events related to artificial intelligence, education, or both fields. After initially showing interest in participating, potential participants received an email invitation along with a letter providing information about the details of the study. This study was performed in the Netherlands, and all participants were active in higher education in said country.

Five separate focus groups were conducted in the period from March 2024 to May 2024 (inclusive). Two focus groups each were conducted for students and teachers, and one was conducted for educational administrators. Each focus group was conducted with the group size of five to eight participants to optimize participant engagement and interaction [17]. The discussions revolved around three dilemmas and ethical considerations surrounding the integration of AI in educational settings. Furthermore, pre-prepared questions guided the discussions, enabling the researchers to gather qualitative data on the ENs and perspectives held by each stakeholder group. Through these questions, particular emphasis was placed on addressing gaps identified through the SLR conducted prior to this study.

The qualitative data collected through the focus groups will be coded deductively with a-priori coding [18] using thematic analysis. The six steps of thematic analysis described by Braun and Clarke will be followed with the software tool Dedoose [20]. The starting coding tree for the thematic analysis will be based on the results of the SLR. If any topics are identified through the thematic analysis that do not fit within this coding tree, they will be added inductively to the coding tree. The ENs and viewpoints elicited from the focus group discussions will serve as valuable inputs for the subsequent stages of the

research project.

The stakeholder consultation aims to gather input from the various stakeholder groups involved in AIED regarding the ethical considerations and challenges associated with using AIED. This input will then be integrated into the development of an ethical framework for AIED. Moreover, the data analysis process will involve mapping the ENs identified by stakeholders onto the main EVs for AIED. This process will also verify the comprehensiveness of the main EVs and supplement the ethical framework with any missing EVs and ENs pertaining to AIED.

2.3. Expert consultation

This research study will engage an expert panel comprising individuals with extensive knowledge and expertise in the domain of AIED. The panel will consist of five to ten experts.

The primary objective of this study is to leverage the collective expertise of the panel in evaluating and prioritizing ENs for an ethical framework for AIED. Based on the methodology described by Sclater, the experts will be tasked with rating the identified ENs on a three-point scale, assessing their criticality to the proposed framework. The ratings will be categorized as follows: 1) Critical, 2) Important, and 3) Less important. This evaluation will facilitate the selection of the most pertinent ENs, which will subsequently be incorporated into the ethical framework.

Furthermore, the study will solicit input from the expert panel regarding the optimal format and presentation of the final ethical framework. This consultation will ensure that the framework is not only theoretically robust but also practical and accessible, facilitating its effective dissemination and adoption within the AIED community and related stakeholders. It is to be noted that the detailed design of this study is still under progress.

The expert consultation will serve to refine and distill the ENs identified through the preceding studies, concentrating on the most critical and essential ones. This refinement process aims to prevent the resulting ethical framework from becoming overly intricate or overwhelming for stakeholders to use effectively. Concurrently, the insights and input garnered from the experts will guide the shaping of the framework into a format that is both usable and comprehensive, ensuring its practical applicability and completeness.

3. Conclusion

In conclusion, this research aims to make a contribution to the field of Artificial Intelligence in Education (AIED) by developing a comprehensive ethical framework to guide ethical usage of AIED. The research follows a rigorous and multifaceted approach, drawing upon various data sources and stakeholder perspectives.

The initial phase of the study will involve a systematic literature review (SLR), which will serve as the foundation for identifying the main ethical values (EVs) and ethical norms (ENs) pertinent to the domain of AIED. This extensive review will consolidate and synthesize the existing knowledge base, providing a theoretical grounding for the subsequent stages of the research.

Building upon this foundation, the second study engage will key stakeholder groups through a series of focus group discussions. This qualitative approach will allow for an in-depth exploration of the perspectives, concerns, and expectations of students, teachers, and educational administrators regarding the ethical implications of AIED. The insights gathered from these discussions will enrich the understanding of the ethical landscape of AIED.

Furthermore, in a third study, we seek the expertise of a panel of domain specialists to validate, rank and refine the proposed ethical framework. The culmination of these efforts would result in a comprehensive ethical framework that addresses the ethical challenges posed by the usage of AIED.

This ethical framework could be used by students, teachers and educational administrators to identify ethical issues in the usage of AIED systems. It could ensure that the usage of AIED does not have any intentional or unintentional adverse effects on their lives. Additionally, it underscores the importance of having a continuous dialogue about the potential ethical issues caused by the rapid evolution of AIED systems. By prioritizing ethical considerations, the educational sector can harness the transformative potential of AI while safeguarding the well-being, privacy, and fundamental rights of all stakeholders involved.

Acknowledgements

I would like to thank my PhD supervisors - Dr. Corrie Urlings, Dr. Giel van Lankveld and Prof. Dr. Roland Klemke - for their unwavering support and guidance, and for always having my back through my PhD trajectory.

References

[1] L. Chen, P. Chen, Z. Lin, Artificial Intelligence in Education: A Review, IEEE Access 8 (2020) 75264–75278. doi:10.1109/ACCESS.2020.2988510.

[2] F. Miao, W. Holmes, R. Huang, H. Zhang, et al., AI and education: A guidance for policymakers, UNESCO Publishing, 2021.

[3] X. Chen, D. Zou, H. Xie, G. Cheng, C. Liu, Two

Decades of Artificial Intelligence in Education, Educational Technology & Society 25 (2022) 28–47.
[4] European Commission and Directorate-General for Communications Networks, Content and Technology, Proposal for a Regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, EC, COM 206 (2021). URL: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206.
[5] W. Holmes, K. Porayska-Pomsta, K. Holstein, E. Sutherland, T. Baker, S. B. Shum, O. C. Santos, M. T. Rodrigo, M. Cukurova, I. I. Bittencourt, K. R. Koedinger, Ethics of AI in Education: Towards a Community-Wide Framework, International Journal of Artificial Intelligence in Education (2021). doi:10.1007/S40593-021-00239-1.
[6] D. Schiff, 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 (2021). doi:10.1007/S40593-021-00270-2/.
[7] P. Blikstein, Y. Zheng, K. Z. Zhou, Ceci n'est pas une école: Discourses of artificial intelligence in education through the lens of semiotic analytics, European Journal of Education 57 (2022) 571–583.
[8] A. Bozkurt, A. Karadeniz, D. Baneres, A. E. Guerrero-Roldán, M. E. Rodríguez, Artificial intelligence and reflections from educational landscape: a review of AI studies in half a century, Sustainability 13 (2021) 800.
[9] P. Lameras, S. Arnab, Power to the teachers: an exploratory review on artificial intelligence in education, Information 13 (2021) 14.
[10] M. Huiling, Research Hotspots and Trends of 'Artificial Intelligence + Education' Abroad in the Last Decade —- Knowledge graph analysis based on 876 articles in the WOS database from 2011 to 2020, in: 2020 The 4th International Conference on E-Society, E-Education and E-Technology, 2020, pp. 132–134.
[11] N. Sclater, Developing a code of practice for learning analytics, Journal of Learning Analytics 3 (2016) 16–42. URL: https://learning-analytics.info/index.php/JLA/article/view/4512. doi:10.18608/jla.2016.31.3.
[12] L. Zhang, K. Fu, X. Liu, Artificial Intelligence in Education: Ethical Issues and its Regulations, in: Proceedings of the 5th International Conference on Big Data and Education, 2022, pp. 1–6.
[13] S. Kagan, Normative ethics, Routledge, 2018.
[14] I. Van de Poel, L. Royakkers, Ethics, technology, and engineering: An introduction, John Wiley & Sons, 2011.
[15] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, International journal of surgery 88 (2021).
[16] N. A. Smuha, Pitfalls and pathways for Trustworthy Artificial Intelligence in education, in: The Ethics of Artificial Intelligence in Education, Routledge, 2022, pp. 113–145.
[17] R. A. Krueger, Focus groups: A practical guide for applied research, Sage publications, 2014.
[18] M. B. Miles, A. M. Huberman, J. Saldana, Qualitative data analysis - international student edition, 4 ed., SAGE Publications, Thousand Oaks, CA, 2019.
[19] V. Braun, V. Clarke, Using thematic analysis in psychology, Qualitative research in psychology 3 (2006) 77–101.
[20] M. Salmona, E. Lieber, D. Kaczynski, Qualitative and mixed methods data analysis using Dedoose: A practical approach for research across the social sciences, Sage Publications, 2019.