



Ethical principles for artificial intelligence in education

Andy Nguyen¹ · Ha Ngan Ngo² · Yvonne Hong³ · Belle Dang¹ · Bich-Phuong Thi Nguyen⁴

Received: 8 February 2022 / Accepted: 31 August 2022 / Published online: 13 October 2022
© The Author(s) 2022

Abstract

The advancement of artificial intelligence in education (AIED) has the potential to transform the educational landscape and influence the role of all involved stakeholders. In recent years, the applications of AIED have been gradually adopted to progress our understanding of students' learning and enhance learning performance and experience. However, the adoption of AIED has led to increasing ethical risks and concerns regarding several aspects such as personal data and learner autonomy. Despite the recent announcement of guidelines for ethical and trustworthy AIED, the debate revolves around the key principles underpinning ethical AIED. This paper aims to explore whether there is a global consensus on ethical AIED by mapping and analyzing international organizations' current policies and guidelines. In this paper, we first introduce the opportunities offered by AI in education and potential ethical issues. Then, thematic analysis was conducted to conceptualize and establish a set of ethical principles by examining and synthesizing relevant ethical policies and guidelines for AIED. We discuss each principle and associated implications for relevant educational stakeholders, including students, teachers, technology developers, policymakers, and institutional decision-makers. The proposed set of ethical principles is expected to serve as a framework to inform and guide educational stakeholders in the development and deployment of ethical and trustworthy AIED as well as catalyze future development of related impact studies in the field.

Keywords Artificial Intelligence · AIED · Ethics · Policies · Privacy

1 Introduction

The application of artificial intelligence (AI) in education has been featured as one of the most pivotal developments of the century (Becker et al., 2018; Seldon with Abidoye, 2018). Despite the rapid growth of AI for education (AIED) and the surge

in its demands under the COVID-19 impacts, little is known about what ethical principles should be in guiding the design, development, and deployment of ethical and trustworthy AI in education. And even if those are addressed, the depth and breadth to which contemporary ethical and regulatory frameworks are able to capture the impacts of AI evolution remain unfolded.

The complexity and “intelligence” of this technology have led to potentially extensive ethical threats that trigger a pressing need for risk-intensive procedures to ensure the quality of delivery. Indeed, a sense of flexibility that acknowledges human values within the developing momentum of AI is vital to fostering sustainable innovations. In the wake of such demand, UNESCO launched global standards for AI ethics which were agreed and signed by its 193 member countries on November 25, 2021. The document, whilst recognizing the “profound and dynamic” influences of AI, also highlights related flourishing dangers to the cultural, social, and ecological diversity (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2021). Notably, it stipulates a universal framework of values for ethics which provides stakeholder-driven guidelines in adopting AI. This historic cross-border agreement marks the globally significant role of ethics in AI; however, it provides a relatively generic framework across disciplines and settings. In fact, for the development and governance of AI technologies, neither *laissez-faire* nor one-size-fits-all approach is adequate and appropriate across contexts. In the literature, ongoing debates regarding ethics of data exploitation in decision making and interventions occur cross-disciplines (Jalal et al., 2021; Farris, 2021; medical care as in Reddy et al., 2020) or human resources management, as in Tambe et al., (2019); sports performance analysis as in Araújo et al., (2021). Recently, researchers and international organizations have specifically examined the ethics of AI in education (Holmes et al., 2021). Despite there being some overlaps and common agreements among these ethical guidelines and reports, no previous study has systematically assessed a global consensus on ethics for AIED.

Our study attempts to fill these gaps by examining and matching ethical guidelines and reports from UNESCO Ethics AI (Ad Hoc Expert Group [AHEG], 2020), UNESCO Education & AI (Miao et al., 2021), Beijing Consensus (UNESCO, 2019), OECD (Organization for Economic Co-operation and Development [OECD], 2021), European Commission (2019), and European Parliament Report AI Education (2021). We sought to prescribe a set of ethical principles for trustworthy AIED based on the thematic analysis results. The establishment of unified ethical principles for AIED gives the research agenda in this domain a new opportunity to meet the demands of a widespread digitalization of education.

This paper is organized as follows. We first introduce a holistic picture of AI in education and present the emerging opportunities. Then, we provide a critical review of the extant literature on ethical issues of AI in education. Next, we present the thematic analysis results of the relevant ethical guidelines and reports for AIED then discuss the implications for associated educational stakeholders. Finally, we conclude by highlighting the significance of ethics in the contemporary discussion of education through which propose several key ethical principles that underpin ethical and trustworthy AIED.

2 Opportunities of artificial intelligence in education

The penetration of AI in every sphere of educational practices has undeniably filtered teachers' and students' personal and professional development with numerous opportunities (Xu & Ouyang, 2021; Ouyang et al., 2022). Existing literature has witnessed a wide diversity of perspectives on the use of AI in education, ranging from the non-teaching aspects (e.g., timetabling, resource allocation, student tracking, provision of information about students to their parents/guardians (reports) to the personalization of teaching and learning (tailored design and marking of assessments, curriculum and AI apps that support learners, or locate changes in learner engagement during foreign language learning (Fahimirad & Kotamjani, 2018; Luckin, 2017; Reiss, 2021; Skinner et al., 2019). Hwang et al., (2020) identified four key roles of AI in education driven by an applications-based perspective that espouses the position of AI as an intelligent tutor, tutee, learning tool/partner, or policy-making advisor.

AIED is seen as an influential tool to empower new paradigms of instruction, technology advancement, and innovations in educational research that are deemed unfeasible in the conventional classroom settings, for instance, the implementation of artificial neural networks, machine learning or CALL (Computer-Assisted Language Learning) in formal, non-formal and informal learning scenarios (Holmes et al., 2019; Hwang et al., 2020). It enables computer-assisted collaborative learning or asynchronous discussion groups, allows cost-wise personalized learning through a navigation system underpinned by algorithms (Nye, 2015), promoted by the use of automated assessment, facial recognition systems, and predictive analytics (Akgun & Greenhow, 2021). Hence, there are growing evidence for the roles of AIED to “foster a transformation of knowledge, cognition, and culture” (Hwang et al., 2020, p.1). However, the implementation of AIED has faced several challenges related to ethical concerns and justification. Although recent attempts have been made to provide ethical guidelines for AIED, there remains the question of a global consensus and standard guidelines for AIED. As the regulation and ethical consensus of these technologies is needed for utilizing their various capabilities in education, this paper sought to offer an integrated overview of ethical guidelines for AIED.

3 Ethical issues of AI in education

Despite its capability to revolutionize education, numerous challenges also linger for researchers and practitioners who are involved in associated activities or systems (Kay & Kummerfeld, 2019) as AIED is, by nature, a “highly technology-dependent and cross-disciplinary field” (Hwang et al., 2020, p.2). At a global level, UNESCO (2019) pinpointed six challenges in achieving sustainable development of AIED: comprehensive public policy, inclusion and equity in AIED, preparing teachers for AI-powered education, preparing AI to understand education, developing quality and inclusive data systems, making research on AIED significant, ensuring ethics and transparency in data collection, use, and dissemination. At the individual level, challenges range from critical societal drawbacks such as systemic bias, discrimination, inequality for marginalized groups of students, and xenophobia (Hwang et al., 2020)

to thorny ethical issues relating to privacy and bias in data collection and processing (Holmes et al., 2021). In fact, the widespread ramifications of AIED have also led to emerging concerns over the negative realities that it brings, such as the widening gaps of inequalities among learners' commercialization of education, or the home-school divide in education (Reiss, 2021). AI may become pervasive in every sense where those involved may be exposed to risks without being aware of them, and the situation can be even intensified under the ongoing impacts of the COVID-19 pandemic (Borenstein & Howard, 2021). Such obstacles essentialize an urgent demand to induct and acquaint teachers and students with the ethical concerns surrounding AIED and how to navigate them.

Furthermore, AIED also carries ethical implications and privacy risks which call for critical attention to differentiate between doing ethical things and doing things ethically (Holmes et al., 2021), or as in the words of Russell and Norvig (2002) "all AI researchers should be concerned with the ethical implications of their work" (p. 1020). Indeed, a proliferation of studies has revealed the emergence of contrasting ethical themes relating to general AI and AIED, most of which are associated with the liability of data across settings, such as in higher education (Zawacki-Richter et al., 2019), K-12 (Holstein et al., 2019), schools (Luckin, 2017), and subjects (Hwang & Tu, 2021). These covered the issues of informed consent, privacy breach, biased data assumption, fairness, accountability, and statistical apophenia. Others also question the impacts of AI-related fields such as surveillance and consent, learner privacy (Sacharidis et al., 2020), identity configuration, user confidentiality, integrity, and inclusiveness (Deshpande et al., 2017). Another stream of discussion has been drawn upon the ethics of data designated for educational use and analytics learning (e.g., Kay & Kummerfeld 2019; Kitto & Knight, 2019; Slade & Prinsloo, 2013). These incorporate the spheres of data interpretation and management, different perspective on the data usage, and the power relation among involved stakeholders such as students, teachers, and the educational objectives (Slade & Prinsloo, 2013). Other ethical issues for AIED include the problems with data collection, restricted availability of data sources, bias and representation, data ownership and control, data autonomy, AIED systems, and human agency (Akgun & Greenhow, 2021; Miao et al., 2021). That said, it is crucial to fully comprehend these values and principals before making ethically and accountability-driven decisions, and being aware of possible, even unexpected outcomes in education.

Although recent work has attempted to establish different ethical frameworks for general AI use (e.g., Ashok et al., 2022), ethical and privacy issues are suggested to be contextualized (Ifenthaler & Schumacher, 2016), hence the prior guidelines established in other disciplines might not be appropriate for education. The contextual approach to the ethical design and use of AIED could play an essential role in addressing the issues of ethical and privacy concerns in education context. Prior research has emphasised the importance of the sociotechnical context configured by educational technology and educations practices in ethical considerations (Kitto & Knight, 2019). The understanding of ethics and privacy from various perspectives could promote the design of ethical and trustworthy AIED and the adoption of such systems. Furthermore, we extended the ethical view from published studies reviewed by Ashok et al., (2022) to the policies and guidelines proposed by the international organizations such

as UNESCO, OECD, and European Union. The consensus assessment of policies and guidelines would inform a comprehensive and integrated instructions for different stakeholders in adopting AIED. This contributes to establishing a common ground and solid foundation for further development and implementation of AIED.

4 Ethical principles for artificial intelligence in education (AIED)

There are continued calls for substantial ethical guidelines and open communications with beneficiaries: educators, students, parents, AI developers, and policymakers (Berendt et al., 2020; Nigam et al., 2021; Hagendorff, 2020) stated that more emphasis is necessary to enforce ethical guidelines for AI systems to better align with societal values. Safeguard measures and human oversight are required to oversee how these AI systems are designed, how they function and evolve. The knowledge of behavioral science, equipped with self-awareness and empathy at the fore, is argued to intrinsically motivate AI developers to develop more trustworthy and responsible AI (Dhanrajani, 2018).

We conducted thematic analysis on relevant ethical guidelines and reports related to AIED found from international organizations, including UNESCO Ethics AI (AHEG, 2020), UNESCO Education & AI (Miao et al., 2021), Beijing Consensus (UNESCO, 2019), OECD (Organization for Economic Co-operation and Development, 2021), European Commission (2019), and European Parliament Report AI Education (2021). The paper focused on identifying and developing a set of main principle themes by using inductive analysis, based on Braun & Clarke (2012)'s thematic analysis process. The analysis consists of initial familiarization with the ethical guidelines and reports. This involved re-reading of reports and noting down patterns, such as similar use of words, points of discussion, and definitions. This is followed by an open coding approach where the terms and definitions are meaningfully categorized, followed by labeling each category with a code. This resulted in a total of 39 codes. Next, these codes were examined and collated into patterns of broader meaning, resulting in 7 themes (i.e. principles). The coding and themes generation process was conducted iteratively, where a researcher-researcher corroboration method was also in place to ensure the reliability and validity (Patton, 2015) of the proposed principles and corresponding code mapping in Table 1.

4.1 Principle of governance and stewardship

A recurring theme across AI policies is the issue of governance and stewardship of AIED (Ashok et al., 2022). For example, the 2021 UNESCO Education & AI 2021 asserted the need to “set up a system wide organizational structure for policy governance and coordination” (Miao et al., 2021, p32). This is further acknowledged in other papers such as OECD (2021, p.4) recommendation for “Principles for responsible stewardship of trustworthy AI”. AIED governance and stewardship declares and manages how AI should be employed in education and relevant mechanisms to assure the compatibility between the role of the technology being deployed and its designed purposes, to optimize educational stakeholders' needs and benefits. AI

principle of governance has been formally defined as “the practice of establishing and implementing policies, procedures and standards for the proper development, use and management of the infosphere.” (Floridi, 2018, p.3). Meanwhile, AI stewardship could be defined as ethics embodied in the careful and responsible management of the design and use of AIED. Although governance and stewardship have been mentioned in most ethical guidelines and policies for AIED, these issues have been surprisingly disregarded from many contemporary ethical debates in the literature (Ashok et al., 2022). While governance refers to “a structure or pattern, stewardship is an activity” (Greer, 2018, p.42). In other words, taking action on issues such as building capacity or developing transparency from a long list of policies can be seen as good stewardship or setting up a better governance. According to OECD Principles for responsible stewardship of trustworthy AI (OECD, 2021), there are five complementary principles relevant to all stakeholders: (i) inclusive growth, sustainable development and well-being; (ii) human-centred values and fairness; (iii) transparency and explainability; (iv) robustness, security and safety; and (v) accountability. While the first and second principles sought to attain the inclusiveness and human-centredness in AIED, the later three OECD principles share several common intersections with data ethics and physical safety in using AIED. Accordingly, we propose that the governance and stewardship of AIED should accomplish all ethical aspects of relevant domains.

Principle of governance and stewardship: The governance and stewardship of AIED should carefully take into account the interdisciplinary and multi-stakeholder perspectives as well as all ethical considerations from relevant domains, including but not limited to data ethics, learning analytics ethics, computational ethics, human rights, and inclusiveness.

The consideration of soft and hard ethics from relevant domains in the governance and stewardship of AIED is critical for the ethical design and use of trustworthy AIED and enhancing its societal implications.

4.2 Principle of transparency and accountability

Data ethics emphasized the need for transparency in data usage in AIED (Larsson & Heintz, 2020). AI tools have been gradually applied quite extensively in education to enhance learning and teaching practices (Wang & Cheng, 2021), but the challenge remains unaddressed regarding the transparency of the data generated. Cope & Kalantzis (2019) highlighted that this ethical principle is essential to teachers and students as data visualization represents learner behavior, and accentuates additional support that educators could provide. It should be noted that the transparency lies in what the data itself is, where it is collected, what it shows, what happens to it, and how it is used (Digital Curation Centre, 2020). These questions could be answered once data ownership, accessibility, and explainability are sustained.

The notion of data ownership, by nature, is a matter of transparency and fairness (Remian, 2019), dealing with who owns and is entitled to the rights to access the personal data of learners. Although technically speaking, consent may often be given to data collectors, whether the data usage intrudes on learners’ privacy has long

Table 1 Ethical Principles for Artificial Intelligence in Education

| Ethical Principles for AI/ED | Codes | Sources | | | | | | |
|----------------------------------|----------------------------|------------------------------|------------------------------|--------------------------|-------------|----------------------------|--|--|
| | | UNESCO Ethic AI 2020 (Draft) | UNESCO Education & AI (2021) | Beijing Consensus (2019) | OECD (2021) | European Commission (2019) | European Parliament Report AI Education (2021) | |
| Governance & Stewardship | General | | | | | | | |
| | Governance & Stewardship | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| | Multistakeholder | ✓ | ✓ | | ✓ | ✓ | | |
| | Interdisciplinary planning | | ✓ | | | | | |
| Transparency & Accountability | International Cooperation | ✓ | | | ✓ | | | |
| | Monitoring & Evaluation | ✓ | ✓ | | ✓ | ✓ | ✓ | |
| | Transparency | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| | Explicability | | | | | ✓ | | |
| | Explainability | ✓ | ✓ | | | ✓ | ✓ | |
| | Responsibility | ✓ | | | | | | |
| Sustainability & Proportionality | Accountability | ✓ | | | | ✓ | | |
| | Auditability | ✓ | ✓ | ✓ | | ✓ | ✓ | |
| | Sustainability | ✓ | ✓ | | ✓ | ✓ | ✓ | |
| | Environment | ✓ | | | ✓ | ✓ | | |
| | Local Alignment | ✓ | ✓ | ✓ | | | | |
| | Proportionality | ✓ | ✓ | ✓ | | ✓ | | |
| Privacy | Economy & Labour | ✓ | ✓ | | ✓ | ✓ | ✓ | |
| | Lifelong learning | | ✓ | ✓ | | | | |
| | Data Privacy | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| | Children's Privacy | | ✓ | | | | ✓ | |
| Security & Safety | Data governance | | | | | ✓ | ✓ | |
| | Safety | ✓ | | | ✓ | ✓ | ✓ | |
| | Robustness | | | ✓ | ✓ | ✓ | ✓ | |
| | Prevention of harm | ✓ | ✓ | ✓ | | ✓ | ✓ | |
| | Security | | | | ✓ | ✓ | ✓ | |

Table 1 (continued)

| Ethical Principles for AI/ED | Codes | Sources | | | | | |
|------------------------------|------------------------|------------------------------|------------------------------|--------------------------|-------------|----------------------------|--|
| | | UNESCO Ethic AI 2020 (Draft) | UNESCO Education & AI (2021) | Beijing Consensus (2019) | OECD (2021) | European Commission (2019) | European Parliament Report AI Education (2021) |
| Inclusiveness | General | | | | | | |
| | Inclusiveness | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Accessibility | | | | | ✓ | ✓ |
| | Diversity | ✓ | ✓ | | | ✓ | ✓ |
| | Integrity of data | | | | | ✓ | ✓ |
| | Non-discriminate Data | ✓ | | | | ✓ | ✓ |
| | Algorithms biases | | ✓ | | | ✓ | ✓ |
| | Fairness | ✓ | ✓ | | ✓ | ✓ | ✓ |
| | Gender equality | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Human Oversight | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Human-centric/centered | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Human Rights | ✓ | | | | ✓ | |
| | Human Dignity | ✓ | | | | ✓ | |
| Human-centered AI/ED | Human (Learner) Agency | | ✓ | ✓ | | ✓ | |
| | Human Autonomy | ✓ | | | ✓ | ✓ | ✓ |
| | Teacher Role | | ✓ | ✓ | | | |
| | | | | | | | |

remained controversial. A valid argument aligned with the integrity of the motive for data collection could be proposed, in which the ownership should be granted to students themselves. Indeed, students are those providing data, thereby having the rights to own and control how data should be used to benefit their own learning (Holmes et al., 2021). Meanwhile, there comes a plausible claim about the rights of institutions to access and use student data since interactions and performance of learners, in essence, are recorded using a structured learning system provided by these educational institutions.

The concept of explainability in AI and data is closely linked to the transparency of the AI system and data generated. Indeed, data should feature the ability to explain some predictions from a technical viewpoint of a particular human. AI explainability underscores the insights in how AI systems functioned and made a decision should be well informed and explicable to the stakeholders, though the explicability relies on their technical expertise and role (Kazim & Koshiyama, 2021; UNESCO, 2019). The opaque nature of AI often poses many challenges for stakeholders to fathom the logic of this “black box” behind its decision-making. For instance, the absence of explainability could result in teachers being unable to use AIED effectively and timely detect the problems related to students’ behavior and learning performance (Remian, 2019). Therefore, this ethical concern centers on the intelligibility of the operation and outcomes of AI educational systems.

Principle of transparency in data and algorithms: The process of collecting, analyzing, and reporting data should be transparent with informed consent and clarity of data ownership, accessibility, and the purposes for how data will be used. The AI algorithms should be explainable and justifiable for specific educational purposes.

The transparency of AIED has been highlighted in several ethical guidelines, including the European Commission’s ethics guidelines for trustworthy AI (2019), European Parliament (2021), UNESCO Education & AI (Miao et al., 2021), Beijing Consensus UNESCO (2019), and OECD’s Principles for responsible stewardship of trustworthy AI (2021). However, the components and descriptions of transparency vary among these reports and guidelines. For instance, while the European Commission (2019, p.18) explained it as “closely linked with the principle of explicability and encompasses transparency of elements relevant to an AI system: the data, the system and the business model”, UNESCO 2020 Draft points to its association “to adequate responsibility and accountability measure” (AHEG, 2020, p.10). Alongside the transparency in data and algorithms, transparency should be of utmost significance to all AIED regulations.

Principle of Transparency in Regulation: The process of establishing, conducting, monitoring, and controlling regulations of AIED should be transparent, traceable, explainable, and communicable in an open and clear manner with clarity of regulatory roles, accessibility, responsibilities, the purposes for how AI will be developed and used, and under which conditions. Additionally, the regulation of AIED should

be transparent in its auditability, and it also links with the next ethical principle of regulatory accountability.

Principle of accountability: The regulation of AIED should explicitly address acknowledgment and responsibility for each stakeholder's actions involved in the design and use of AIED, including auditability, minimization, and reporting of negative side effects, trade-offs, and compensation.

The accountability of AIED relates to the concept of “responsible AI” that features the ethical practice of designing, developing, and implementing AI with good intentions to empower relevant stakeholders and society fairly. Though ‘responsible AI’ has become increasingly popular, the terms accountability and responsibility are rarely defined (Jobin et al., 2019). Nevertheless, it is commonly referred to as acting with integrity and clearly determining the attribution of responsibility and legal liability with careful consideration of potentially harmful factors. AI has been questioned over whether it should be held accountable in a human-like manner, or whether humans should always be the sole actors responsible for AI as technological artifacts. In conjunction with human-centered AIED that encouraged human oversight over AI, we recommended the latter case that educational stakeholders should always be the responsible ones for AIED. Furthermore, some AI policies have highlighted that regulation of AIED should step beyond the scope of individual and organizational accountability to also consider sustainability and proportionality (AHEG, 2020).

4.3 Principle of sustainability and proportionality

Similar to other technology advances, the development and deployment of AI should also take into account environmental concerns to the extent that is referenced (AHEG, 2020; OECD, 2021). Particularly, sustainability calls for the design, development, and use of AIED to consider optimizing energy efficiency and minimizing its ecological footprint (European Commission, 2019). Accordingly, regulations of AIED are required to create policies ensuring these considerations are accomplished throughout the processes of developing and deploying AIED. Moreover, regulation of AIED must consider other sustainable domains, including economic and societal aspects such as employability, culture, and politics (European Parliament, 2021).

Principle of sustainability and proportionality: AIED must be designed, developed, and used in a justifiable way that they would not disrupt the environment, world economy, and society, such as the labor market, culture, and politics.

For instance, regulation of AIED should consider ensuring policies supporting accountability of potential job losses and to leverage challenges as an opportunity for innovation (UNESCO, 2019). Careful deliberations of sustainability and proportionality will make AIED more approachable and beneficial to all.

4.4 Principle of privacy

Personal privacy also emerged as a critical ethical concern in the implementation of AIED. Privacy, by nature, could be defined as “the right to be left alone”, which underscores the right of having personal information being protected (Muller, 2020). This digital revolution in education, particularly the use of AI and learning analytics in the field of education, entails a massive amount of personal data generated, captured, and analyzed to optimize learning experiences (Tzimas & Demetriadis, 2021; Pardo & Siemens, 2014). The personal data of teachers and learners may run the risk of privacy breaches. For instance, in respect of agent-based personalized education, personal information of learning performance accumulated in the past could be utilized for future prediction. However, this is considered against the will of many students (Li, 2007).

To protect and support the right of learners’ privacy and social well-being while learning in the context of increasingly knowledgeable machines and computer agents, AIED developers need to assess the views of teachers and students to decide how AI should be deployed in the classroom (Miao et al., 2021). For instance, an ethical concern may arise from a real-time facial expression recognition system used to predict the affective state (e.g. Jian-Ming Sun et al., 2008) or attendance of the learners without their consent (e.g. Pattnaik & Mohanty 2020). Developers and educators should embed transparency and visibility to AIED-related threats while explaining potential ramifications to students’ learning, careers, and social lives. The objective is to cultivate trust among learners and provide them with insights to leverage their skills across contexts while maintaining control of their respective data and digital identities (Jobin et al., 2019).

Principle of privacy: AIED must ensure well-informed consent from the user and maintain the confidentiality of the users’ information, both when they provide information and when the system collects information about them.

In most cases, when AIED tools are used to engage users in a particular learning activity, users are assumed to give consent, by which they would agree on terms of use of technology and how their personal data is collected, managed, and processed. Aligned with the principle of transparency and accountability, consent must be well informed as a pragmatic approach to building trust among students since the consent demonstrates their ease with the use of data by teachers to enhance their own learning performance (Li et al., 2021; Sedenberg & Hoffmann, 2016) also highlighted the significance of this consent to show respect towards students and reinforce their autonomy and freedom of choice. Once data is garnered, questions arise about how data management works, where and how long their personal information should be stored, and to whom the rights of accessibility should be granted (Corrin et al., 2019).

4.5 Principle of Security and Safety

One of the main functions of educational learning systems is to collect data of users, from which predictions about the learning behaviors and performance of users will

be made. However, it is inevitable to envisage a scenario when the data is probably manipulated or corrupted by another party, or even worse, by cybercriminals.

Principle of Security: AIED should be designed and implemented in a manner that ensures the solution is robust enough to safeguard and protect data effectively from cybercrimes, data breaches and corruption threats, ensuring the privacy and security of sensitive information.

The concept of incorruptibility in AIED traces its root from incorruptibility in AI, or robustness against malicious manipulation by external factors. Bostrom & Yudkowsky (2014) pointed out that AI systems must be “robust against human adversaries deliberately searching for exploitable flaws in the algorithm” (p. 317). Therefore, it can be stated that the incorruptible nature and integrity of the data go hand in hand with data security. It is essential to protect the personal data of stakeholders, including students, teachers, and schools, to prevent any misuse or violation. The protection of data privacy and security is even more essential in the current context of normalizing virtual learning, and it requires concerted effort and the self-awareness of all the stakeholders.

Whereas learning analytics are governed by data ethics, many AIs are forms of intelligence expressed by some artifact (Bryson & Theodorou, 2019) that interacts with humans at various levels, such as robots and self-driving cars (Manoharan, 2019; O’Sullivan et al., 2019). This raises a universe of technical safety concerns regarding AI operation throughout its lifecycle in normal use, especially in harsh conditions or where other agents (both human and artificial) can interfere with the system.

Principle of Safety: AIED systems to be designed, developed, and deployed in a risk-management approach so that users are protected from unintended and unexpected harm, and that fatalities are mitigated.

As a result, it is pivotal that AIED developers take great care to design, train, pilot test, and validate the safety of AI systems (Leslie, 2019). Multistakeholder groups, including product developers, educators, and public authorities, should establish appropriate oversight, assessment, and due diligence mechanisms to ensure accountability and robustness throughout the AI lifecycle (AHEG, 2020). This group should produce detailed guidelines and ensure that AI users (educators and learners) receive adequate training to operate the system safely within the defined environment.

4.6 Principle of inclusiveness

Previous ethical discourse suggested that AI systems should contribute to global justice and be equally accessible to all (European Commission, 2018). Accessibility is vital to allow society to gain significant benefits from these systems. The exclusion of any individual is a violation of human rights. It is, hence, paramount that accessibility entails affordability, user-friendly designs catering to individuals of different demographics, cultures, and particularly those with disabilities (Kazim & Koshi-

yama, 2021). As highlighted in the European Commission Report 2021, inclusion and fairness of access to AI-powered education stress the basic needs and availability for internet coverage, followed by next-generation digital infrastructure.

Principle of Inclusiveness in Accessibility: AIED design, development, and deployment must take into account the infrastructure, equipment, skills, and societal acceptance that will accommodate a wide range of individuals in the intended region, allowing equitable access and use of AIED.

The current digital gap evidently widens after COVID-19, where countries with poor infrastructure hamper their aspirations of thriving in digitalization (Palomares et al., 2021). Furthermore, the fundamental lack of access to technologies, such as students from socially disadvantaged backgrounds not owning personal digital devices (Sá et al., 2021), calls for collective discussions with all educational stakeholders on the aspects of inclusion in AIED (i.e. addressing the lack of opportunities, resource sharing efforts to counter areas suffering from deprivation of learning resources, and knocking down discriminatory structures) to reduce educational inequities (Office of the High Commissioner for Human Rights, 2019).

Another aspect of inclusiveness is non-discrimination or unbiased AI algorithms. Quality education is fundamental in fostering a flourishing society, where all learners are viewed equally regardless of their gender, race, beliefs, sexual orientation, and any other conditions or circumstances (Palomares et al., 2021). AIED design requires careful considerations to avoid discrimination against certain groups, as AIED relies on and will only be as good as its trained data. Hence, it is crucial that AI developers take precautions by training the AIED with comprehensive and diverse data to reduce instances where the AIED would manifest a particular bias (Hogenhout, 2021) and violate the non-maleficence principle.

Principle of Inclusiveness in Data and Algorithms: AIED design, development, and deployment must apply non-discrimination and unbiased data and algorithms to ensure fairness and equality among different groups of beneficiaries.

Data quality plays a crucial role in determining whether AIED could make valid and unbiased decisions since bias manifests itself in the AIED system with the biased training data (Borgesius, 2018; Digital Curation Centre, 2020). Several aspects of biased data relating to gender, race, ethnicity, and special learning needs. An illuminating example in language education technology is given by West-Smith et al., (2018) that input data in the form of rubric writing and scoring system may place a constraint on the task choice and writing styles of students. Thus, there is a need for bias-free data in AIED to avoid biased algorithms.

4.7 Principle of human-centered AIED

In recognition of autonomy as a modern moral and political value (Calvo et al., 2020), the development and regulation of AIED need to adopt a human-centric approach that safeguards and empowers human autonomy. This principle emphasizes the impor-

tance of supporting learners in developing their own potential (Miao et al., 2021; UNESCO, 2019).

Principle of Human-Centred AIED: The goal of AIED should be to complement and enhance human cognitive, social, and cultural capabilities while preserving meaningful opportunities for freedom of choice, securing human control over AI-based work processes.

Human autonomy, according to Deci and Ryan (2020), refers to the capacity to live one's life according to one's own motivation that is not the result of deception or manipulation. AI assistants today serve a variety of functions, generally intending to provide and assist individuals with some recommendations. In a sense, these can be considered external factors that affect an individual's cognitive bias and emotions undermining or manipulating one's intrinsic motivation (Vesnic-Alujevic et al., 2020). The design and operation of AI must thus, avoid misleading information, compromising users' autonomy in developing independent thoughts, or negatively affecting users' emotions and social well-being.

Research and development on AIED must avoid algorithms and wordings that serve as computational propaganda (Brundage et al., 2018; Nobre, 2020) in the form of automated feedback, learning assessment, and suggestions. This is particularly pertinent in an educational context, where many users are children and young people, constituting a vulnerable group that deserves special care and protection (European Parliament, 2021). There should be training programs supporting educators to gain the required skills to implement AIED. They should be able to adapt, filter or reduce automation that might coerce and manipulate learners' thinking, impeding rather than supporting their motivation and identity development.

The focus of the previous dimension may be defined as the autonomy of will (Caughey et al., 2009) or positive freedom, referring to the capacity to develop one's independent wishes and intrinsic motivations. Nevertheless, autonomy also underpins the autonomy of action (Möller, 2009), which refers to the ability to act on preferences without external restrictions. The AIED system relies on vast amounts of data to make predictions, which, in many cases, results in undesirable deductions of options to prevent users from engaging in or performing actions that the system views as errors (Bryson & Theodorou, 2019). Facebook's decision to change its algorithm to prevent fake news and use fake IDs is one example. These interventions from AI, regardless of best intentions, potentially limit individual freedom of expression (an identity) or perform certain actions. Therefore, Fagan & Levmore (2019) suggested that humans ought to remain in the center of AI design and implementation, to be the ones presumably deciding the goals of AI and have the power to overrule machine decisions.

Of all AIED, a tool for assessing and providing guidance for students, predominantly referred to as an "intelligent tutoring system", is the longest researched and most common application (Miao et al., 2021). The system mapped out learning materials and activities based upon experts' knowledge of the subject and cognitive sciences, as well as student misconceptions and success. With increasing automated decisions and shortcut suggestions made by machines, it is likely that AIED

will reduce learners' interaction with others and their ability to cultivate individual resourcefulness, metacognition, self-regulation, and independent thought. One of AIED's main ethical concerns is the possibility of undermining learner agency and pertaining to breaching autonomy of action.

To ensure a human-centric AIED that emphasizes the learner agency, researchers, developers, and practitioners must adopt an interdisciplinary approach to developing negotiation-based adaptive learning systems that emphasize but are not limited to transversal competencies (European Parliament, 2021). AIED should allow learners the power to negotiate the type and frequency of received support, scaffolding of not only knowledge but also metacognition and self-regulation skills (Chou et al., 2018; Daradoumis & Arguedas, 2020). Governments and educators should be aware of the AI literacy skills crucial for effective human-machine collaboration to develop and integrate the appropriate curriculum into education practices. As a result, not only will students and teachers remain in control and at the center of AI implementation, but humans and machines will also collaborate for improved educational outcomes rather than using AI to usurp humans (Bryson & Theodorou, 2019).

5 Final remarks and future directions

The education system faces a paradox of artificial intelligence. Though regarded as vital for AI generation of high-quality educational outcomes, AIED and related large-scale collection and analysis of personal data about learners are of considerable concern to human-rights advocates. This paper contributes to the discussions of the benefits of AI in education, and at the same time, raises concerns for the adverse impacts on fundamental issues surrounding human rights. The intricacy of AI necessitates a holistic and applicable set of ethical principles for AI in the educational context. By systematically analyzing well-documented general AI principles, we propose a set of ethical tenets for AIED as a starting point to engage and spark further debates on the robustness of these guidelines, followed by actionable and shared policies to ensure the AIED systems developed are essentially ethical by design. The proposed set of ethical principles should be considered when developing and implementing ethical and trustworthy AI systems for education.

Nevertheless, given the growing interest of AIED in a post-Covid era, it is foreseeable that this debate concerning will continually evolve and move forward in the long run. A natural progression to be witnessed within the literature, namely a precise mechanism of ethical principles in AIED, remains to be elucidated. Indeed, while educational practitioners and AI developers have the best intentions of developing and implementing AI to improve education, the guiding ethical principles for AIED are yet to be set in stone. Furthermore, the education system has been confronting a paradox of applications of artificial intelligence technology across teaching and learning contexts. Despite existing theoretical frameworks investigating ethics of AI in general, no universal consensus has been reached on the best ethical theory in general, with moderate attention given to a practical set of ethical standards in the field of education in particular. Additionally, though regarded as vital for AI generation of high-quality educational outcomes, AIED and related large-scale collection and

analysis of personal data about learners are of considerable concern to human-rights advocates. Such challenges call for greater attention in effectively and appropriately addressing associated ethical dilemmas. Given the interdisciplinary nature of AI, this is anticipated to be an arduous task to achieve since ethical principles are basically derived from human being judgement which can be largely abstract-driven, frequently inextricably intertwined with subjective interpretations. The engagement of diverse stakeholders in any educational discourse further impedes ethical principles from being widely applied either in a formal or deductive manner. Hence, based on our findings as preliminary ground, future scholarship is encouraged to extend the focus of inquiry to the implementation stage, where the issues of accessibility assurance, bias and equity in adopting AIED, or developmental and neurological influences of AIED to vulnerable groups such as young children and handicapped would be another interesting sphere that deserves continued exploration. Considerably more work will need to be done to establish and validate a common understanding and standards on ethics in AIED. A natural progression of this work is to publish a website about this set of ethical principles for AIED in order to capitalize feedback and improvement suggestions about the use of this framework. Furthermore, an automatic method by text analysis could be conducted in further work to provide complementing findings. Last but not least, embedding the principles of AI ethics in education, and also ethics issues such as responsibility, inclusion, fairness, security and explainability in conducting educational research will not only mitigate emerging societal abuses rooted from algorithmic injustice, but also bear instrumental implications to the landscape of AI governance and policy making for the long-term development of significant industries. Among one of the first papers to spark on the applicability of a set of ethical guidelines and practice standards for artificial intelligence in education, it is also our expectation that this will be a fruitful step towards guiding future educators and learners to exercise, if not instill, stronger accountability and responsibility in adopting AI and the technology that they employ for their teaching and learning in the future.

Overall, this paper contributes to the ongoing discussions of the benefits of AI in education, and at the same time, raises concerns for the adverse impacts on fundamental issues surrounding human rights. The intricacy of AI necessitates a holistic and applicable set of ethical principles for AI in the educational context. By systematically analyzing well-documented general AI principles, we propose a set of ethical tenets for AIED as a starting point to engage and spark further debates on the robustness of these guidelines, followed by actionable and shared policies to ensure the AIED systems developed are essentially ethical by design.

Funding Open Access funding provided by University of Oulu including Oulu University Hospital. This work was funded in part by Finnish Academy project no. 350249

Data Availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors have NO conflict of interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ad Hoc Expert Group (2020). *Outcome document: First draft of the recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000373434>
- Akgun, S., & Greenhow, C. (2021). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 1–10. <https://doi.org/10.1007/s43681-021-00096-7>
- Araújo, D., Couceiro, M., Seifert, L., Sarmento, H., & Davids, K. (2021). *Artificial Intelligence in Sport Performance Analysis*. Routledge. <https://doi.org/10.4324/9781003163589>
- Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management*, 62, 102433. <https://doi.org/10.1016/j.ijinfomgt.2021.102433>
- Becker, S. A., Brown, M., Dahlstrom, E., Davis, A., DePaul, K., Diaz, V., & Pomerantz, J. (2018). *NMC Horizon Report: 2018 Higher Education Edition*. Educause. <https://library.educause.edu/~media/files/library/2018/8/2018horizonreport.pdf>
- Berendt, B., Littlejohn, A., & Blakemore, M. (2020). AI in education: learner choice and fundamental rights. *Learning Media and Technology*, 45(3), 312–324. <https://doi.org/10.1080/17439884.2020.1786399>
- 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>
- Borgesius, F. Z. (2018). *Discrimination, artificial intelligence and algorithmic decision-making*. Strasbourg: Council of Europe. <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>
- Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish, & W. Ramsey (Eds.), *The Cambridge Handbook of Artificial Intelligence* (pp. 316–334). Cambridge University Press. <https://doi.org/10.1017/CBO9781139046855.020>
- Braun, V., & Clarke, V. (2012). Thematic Analysis. In APA Handbook of Research Methods in Psychology: Vol 2 (pp. 57–71). American Psychological Association. <https://doi.org/10.1037/13620-004>
- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., hEigeartaigh, S., Beard, S., Belfield, H., Farquhar, S., & Amodei, D. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *ArXiv:1802.07228 [Cs]*. <http://arxiv.org/abs/1802.07228>
- Bryson, J. J., & Theodorou, A. (2019). How society can maintain human-centric artificial intelligence. In *Human-centered digitalization and services* (pp. 305–323). Springer. https://doi.org/10.1007/978-981-13-7725-9_16
- Calvo, R. A., Peters, D., Vold, K., & Ryan, R. M. (2020). Supporting human autonomy in AI systems: A framework for ethical enquiry. In *Ethics of Digital Well-Being* (pp. 31–54). Springer. https://doi.org/10.1007/978-3-030-50585-1_2
- Caughey, D., Cohon, A., & Chatfield, S. (2009). Defining, measuring, and modeling bureaucratic autonomy. *Annual Meeting of the Midwest Political Science Association, Chicago*, 2
- Chou, C. Y., Lai, K. R., Chao, P. Y., Tseng, S. F., & Liao, T. Y. (2018). A negotiation-based adaptive learning system for regulating help-seeking behaviors. *Computers & Education*, 126, 115–128. <https://doi.org/10.1016/j.compedu.2018.07.010>
- Cope, B., & Kalantzis, M. (2019). Education 2.0: Artificial intelligence and the end of the test. *Beijing International Review of Education*, 1, 528–543. <https://doi.org/10.1163/25902539-00102009>

- Corrin, L., Kennedy, G., French, S., Buckingham Shum, S., Kitto, K., Pardo, A., West, D., Mirriahi, N., & Colvin, C. (2019). *The Ethics of Learning Analytics in Australian Higher Education: A Discussion Paper*. <https://melbournecshe.unimelb.edu.au/research/research-projects/edutech/the-ethical-use-of-learning-analytics>
- Daradoumis, T., & Arguedas, M. (2020). Cultivating students' reflective learning in metacognitive activities through an affective pedagogical agent. *Educational Technology and Society*, 23(2), 19–31. <https://www.jstor.org/stable/26921131>
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Deshpande, M., & Rao, V. (2017, December). Depression detection using emotion artificial intelligence. In *2017 International Conference on Intelligent Sustainable Systems (iciss)* (pp. 858–862). IEEE. <https://doi.org/10.1109/ISSI.2017.8389299>
- Dhanrajani, S. (2018). 3 Ways To Human Centric AI. <https://www.forbes.com/sites/cognitiveworld/2018/12/12/3-ways-to-human-centric-ai/?sh=495e42804a38>
- Digital Curation Centre, The University of Edinburgh (2020). *The Role of Data in AI: Report for the Data Governance Working Group of the Global Partnership of AI*. <https://www.research.ed.ac.uk/en/publications/the-role-of-data-in-ai>
- European Commission (2018). *Statement on artificial intelligence, robotics and 'autonomous' systems*. European Union Publications Office. <https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-78120382>
- European Commission (2019). *The European Commission's high-level expert group on artificial intelligence: Ethics guidelines for trustworthy AI*. European Union Publications Office. <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>
- European Parliament (2021). *Report on artificial intelligence in education, culture and the audiovisual sector (2020/2017(INI))*. Committee on Culture and Education. https://www.europarl.europa.eu/doceo/document/A-9-2021-0127_EN.html
- Fagan, F., & Levmore, S. (2019). The impact of artificial intelligence on rules, standards, and judicial discretion. *S Cal L Rev*, 93, 1. <https://doi.org/10.2139/ssrn.3362563>
- Fahimirad, M., & Kotamjani, S. S. (2018). A review on application of artificial intelligence in teaching and learning in educational contexts. *International Journal of Learning and Development*, 8(4), 106–118. <https://doi.org/10.5296/ijld.v8i4.14057>
- Farris, A. B., Vizcarra, J., Amgad, M., Cooper, L. A., Gutman, D., & Hogan, J. (2021). Artificial intelligence and algorithmic computational pathology: an introduction with renal allograft examples. *Histopathology*, 78(6), 791–804. <https://doi.org/10.1111/his.14304>
- Floridi, L. (2018). Soft Ethics and the Governance of the Digital. *Philosophy & Technology*, 31(1), 1–8. <https://doi.org/10.1007/s13347-018-0303-9>
- Greer, S. L. (2018). Organization and governance: Stewardship and governance in health systems. *Health Care Systems and Policies*. New York, NY: Health Services Research. Springer. https://doi.org/10.1007/978-1-4614-6419-8_22-1
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99–120. <https://doi.org/10.1007/s11023-020-09517-8>
- Hogenhout, L. (2021). *Unite Paper | A Framework for Ethical AI at the United Nations*. <https://unite.un.org/news/unite-paper-framework-ethical-ai-united-nations>
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2021). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-021-00239-1>
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for complementarity: Teacher and student needs for orchestration support in ai-enhanced classrooms. In *International Conference on Artificial Intelligence in Education* (pp. 157–171). Springer, Cham. https://doi.org/10.1007/978-3-030-23204-7_14
- Hwang, G. J., & Tu, Y. F. (2021). Roles and Research Trends of Artificial Intelligence in Mathematics Education: A Bibliometric Mapping Analysis and Systematic Review. *Mathematics*, 9(6), 584. <https://doi.org/10.3390/math9060584>
- Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1, 100001. <https://doi.org/10.1016/j.caeai.2020.100001>

- Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development*, 64(5), 923–938. <https://doi.org/10.1007/s11423-016-9477-y>
- Jalal, S., Parker, W., Ferguson, D., & Nicolaou, S. (2021). Exploring the role of artificial intelligence in an emergency and trauma radiology department. *Canadian Association of Radiologists Journal*, 72(1), 167–174. <https://doi.org/10.1177/0846537120918338>
- Sun, J. M., Pei, X. S., & Shi-Sheng, Z. (2008). Facial emotion recognition in modern distant education system using SVM. *2008 International Conference on Machine Learning and Cybernetics*, 3545–3548. <https://doi.org/10.1109/ICMLC.2008.4621018>
- Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: The global landscape of ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kay, J., & Kummerfeld, B. (2019). From data to personal user models for life-long, life-wide learners. *British Journal of Educational Technology*, 50(6), 2871–2884. <https://doi.org/10.1111/bjet.12878>
- Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, 2(9), 100314. <https://doi.org/10.1016/j.patter.2021.100314>
- Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology*, 50(6), 2855–2870. <https://doi.org/10.1111/bjet.12868>
- Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, 9(2), 1–16. <https://doi.org/10.14763/2020.2.1469>
- Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. <https://doi.org/10.5281/zenodo.3240529>
- Li, W., Sun, K., Schaub, F., & Brooks, C. (2021). Disparities in students' propensity to consent to learning analytics. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-021-00254-2>
- Li, X. (2007). Intelligent agent-supported online education. *Decision Sciences Journal of Innovative Education*, 5(2), 311–331. <https://doi.org/10.1111/j.1540-4609.2007.00143.x>
- Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour*, 1(3), 1–3. <https://doi.org/10.1038/s41562-016-0028>
- Manoharan, S. (2019). An improved safety algorithm for artificial intelligence enabled processors in self driving cars. *Journal of Artificial Intelligence*, 1(02), 95–104. <https://doi.org/10.36548/jaia.2019.2.005>
- Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: Guidance for policy-makers*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000376709>
- Möller, K. (2009). Two Conceptions of Positive Liberty: Towards an Autonomy-based Theory of Constitutional Rights. *Oxford Journal of Legal Studies*, 29(4), 757–786. <https://doi.org/10.1093/ojls/gqp029>
- Müller, V. C. (2020). Ethics of Artificial Intelligence and Robotics. <https://plato.stanford.edu/entries/ethics-ai/>
- Nigam, A., Pasricha, R., Singh, T., & Churi, P. (2021). A systematic review on ai-based proctoring systems: Past, present and future. *Education and Information Technologies*, 26(5), 6421–6445. <https://doi.org/10.1007/s10639-021-10597-x>
- Nobre, G. (2020). *Artificial Intelligence (AI) in communications: Journalism, public relations, advertising, and propaganda*. <https://doi.org/10.13140/RG.2.2.33598.31040>
- Nye, B. D. (2015). Intelligent tutoring systems by and for the developing world: A review of trends and approaches for educational technology in a global context. *International Journal of Artificial Intelligence in Education*, 25(2), 177–203. <https://doi.org/10.1007/s40593-014-0028-6>
- O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., Holzinger, K., Holzinger, A., Sajid, M. I., & Ashrafian, H. (2019). Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 15(1), e1968. <https://doi.org/10.1002/rcs.1968>

- Office of the High Commissioner for Human Rights (2019). *Transforming our world: the 2030 Agenda for Sustainable Development*. <https://sdgs.un.org/2030agenda>
- Organization for Economic Co-operation and Development (2021). *OECD Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, 1–33. <https://doi.org/10.1007/s10639-022-10925-9>
- Palomares, I., Martínez-Cámara, E., Montes, R., García-Moral, P., Chiachio, M., Chiachio, J., & Herrera, F. (2021). A panoramic view and swot analysis of artificial intelligence for achieving the sustainable development goals by 2030: progress and prospects. *Applied Intelligence*, 1–31. <https://doi.org/10.1007/s10489-021-02264-y>
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. <https://doi.org/10.1111/bjet.12152>
- Patton, M. Q. (2015). *Qualitative research & evaluation methods: integrating theory and practice* (4th ed.). Los Angeles: SAGE
- Pattnaik, P., & Mohanty, K. K. (2020). AI-Based Techniques for Real-Time Face Recognition-based Attendance System- A comparative Study. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 1034–1039. <https://doi.org/10.1109/ICECA49313.2020.9297643>
- Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, 27(3), 491–497. <https://doi.org/10.1093/jamia/ocz192>
- Reiss, M. J. (2021). The use of AI in education: Practicalities and ethical considerations. *London Review of Education*, 19(1), 5, 1–14. <https://doi.org/10.14324/LRE.19.1.05>
- Remian, D. (2019). *Augmenting education: Ethical considerations for incorporating artificial intelligence in education*. ScholarWorks at UMass Boston. University of Massachusetts Boston. https://scholarworks.umb.edu/instruction_capstone/52
- Russel, S. J., & Norvig, P. R. (2002). *Artificial Intelligence: A modern approach* (2nd Ed.) Prentice Hall Upper Saddle River, NJ, USA
- Sá, M. J., Santos, A. I., Serpa, S., & Ferreira, M., C (2021). Digitainability—Digital Competences Post-COVID-19 for a Sustainable Society. *Sustainability*, 13(17), 9564. <https://doi.org/10.3390/su13179564>
- Sacharidis, D., Mukamakuza, C. P., & Werthner, H. (2020). Fairness and diversity in social-based recommender systems. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 83–88). <https://doi.org/10.1145/3386392.3397603>
- Sedenberg, E., & Hoffmann, A. L. (2016). *Recovering the history of informed consent for data science and internet industry research ethics*. <https://arxiv.org/abs/1609.03266>
- Seldon, A., & Abidoye, O. (2018). *The fourth education revolution*. Legend Press Ltd.
- Skinner, G., & Walmsley, T. (2019, February). Artificial intelligence and deep learning in video games a brief review. In *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)* (pp. 404–408). IEEE. <https://doi.org/10.1109/CCOMS.2019.8821783>
- Slade, S., & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*, 57(10), <https://doi.org/10.1177/0002764213479366>. 1510 – 1529
- Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15–42. <https://doi.org/10.1177/0008125619867910>
- Tzimas, D., & Demetriadis, S. (2021). Ethical issues in learning analytics: a review of the field. *Education Technology Research Development*, 69, 1101–1133. <https://doi.org/10.1007/s11423-021-09977-4>
- United Nations Educational, Scientific and Cultural Organization, & Organization, C. (2021). *Recommendation on the Ethics of Artificial Intelligence*. United Nations Educational. <https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14>
- United Nations Educational, Scientific and Cultural Organization (2019). Beijing Consensus on artificial intelligence and education. *Outcome document of the International Conference on Artificial Intelligence and Education, Planning Education in the AI Era: Lead the Leap, Beijing, 2019*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000368303>

- Vesnic-Alujevic, L., Nascimento, S., & Pólvara, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, 44(6), 101961. <https://doi.org/10.1016/j.telpol.2020.101961>
- Wang, T., & Cheng, E. C. K. (2021). An investigation of barriers to Hong Kong K-12 schools incorporating Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*, 2, 100031. <https://doi.org/10.1016/j.caeai.2021.100031>
- West-Smith, P., Butler, S., & Mayfield, E. (2018). Trustworthy automated essay scoring without explicit construct validity. In *Proceedings of the AAAI Spring Symposium on AI and Society: Ethics, Safety and Trustworthiness in Intelligent Agents*. <https://help.turnitin.com/Resources/RA%20Curriculum%20Resources/Research/Revision%20Assistant%20Validity%20AAAI%202018.pdf>
- Xu, W., & Ouyang, F. (2021). A systematic review of AI role in the educational system based on a proposed conceptual framework. *Education and Information Technologies*, 1–29. <https://doi.org/10.1007/s10639-021-10774-y>
- Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1–27. <https://doi.org/10.1186/s41239-019-0171-0>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Andy Nguyen¹ · Ha Ngan Ngo² · Yvonne Hong³ · Belle Dang¹ · Bich-Phuong Thi Nguyen⁴

✉ Andy Nguyen
andy.nguyen@oulu.fi

¹ Learning & Educational Technology Research Unit (LET), University of Oulu, Oulu, Finland

² Faculty of Education, Victoria University of Wellington, Wellington, New Zealand

³ School of Information Management, Victoria University of Wellington, Wellington, New Zealand

⁴ Faculty of English Language Teacher Education, VNU University of Languages and International Studies, Hanoi, Vietnam