International Journal of Artificial Intelligence in Education (2022) 32:504–526
https://doi.org/10.1007/s40593-021-00239-1

ARTICLE

Ethics of AI in Education: Towards a Community-Wide Framework

Wayne Holmes[1] · Kaska Porayska-Pomsta[1] · Ken Holstein[2] · Emma Sutherland[3] · Toby Baker[3] · Simon Buckingham Shum[4] · Olga C. Santos[5] · Mercedes T. Rodrigo[6] · Mutlu Cukurova[1] · Ig Ibert Bittencourt[7] · Kenneth R. Koedinger[2]

Accepted: 26 January 2021 / Published online: 9 April 2021

Abstract

While Artificial Intelligence in Education (AIED) research has at its core the desire to support student learning, experience from other AI domains suggest that such ethical intentions are not by themselves sufficient. There is also the need to consider explicitly issues such as fairness, accountability, transparency, bias, autonomy, agency, and inclusion. At a more general level, there is also a need to differentiate between *doing ethical things* and *doing things ethically*, to understand and to make pedagogical choices that are ethical, and to account for the ever-present possibility of unintended consequences. However, addressing these and related questions is far from trivial. As a first step towards addressing this critical gap, we invited 60 of the AIED community's leading researchers to respond to a survey of questions about ethics and the application of AI in educational contexts. In this paper, we first introduce issues around the ethics of AI in education. Next, we summarise the contributions of the 17 respondents, and discuss the complex issues that they raised. Specific outcomes include the recognition that most AIED researchers are not trained to tackle the emerging ethical questions. A well-designed framework for engaging with ethics of AIED that combined a multidisciplinary approach and a set of robust guidelines seems vital in this context.

Keywords Artificial intelligence in education · Ethics · Fairness · Agency · Pedagogy · Human cognition

Introduction

It is almost certainly the case that all members of the Artificial Intelligence in Education (AIED) research community are motivated by ethical concerns, such as improving

✉ Wayne Holmes
wayne.holmes@ucl.ac.uk

Extended author information available on the last page of the article

students' learning outcomes and their lifelong opportunities. However, as has been seen in other domains of AI application, ethical intentions are not by themselves sufficient, as good intentions do not always result in ethical designs or ethical deployments (e.g., Dastin 2018; Reich and Ito 2017; Whittaker et al. 2018). Significant attention is required to understand what it means to be ethical *specifically in the context of AIED*. The educational contexts which AIED technologies aspire to enhance highlight the need to differentiate between *doing ethical things* and *doing things ethically*, to understand and to make pedagogical choices that are ethical, to account for the ever-present possibility of unintended consequences, along with many other considerations. However, addressing these and related questions is far from trivial, not least because it remains true that "*no framework has been devised, no guidelines have been agreed, no policies have been developed, and no regulations have been enacted to address the specific ethical issues raised by the use of AI in education"* (Holmes et al. 2018, p. 552).

As a first step towards addressing this critical gap, we invited 60 of the AIED community's leading researchers (operationalised mainly by citations and/or reputation) to respond to a survey of questions about ethics and the application of AI in educational contexts.

The 17 respondents (in alphabetical order by surname) were: Vincent Aleven (CMU, USA), Ig Ibert Bittencourt (UFAL, Brazil), Jesus Boticario (UNED, Spain), Simon Buckingham Shum (UTS, Australia), Mutlu Cukurova (UCL, UK), Ben du Boulay (University of Sussex, UK), Janice Gobert (Rutgers, USA), Ken Holstein (CMU, USA), Judy Kay (University of Sydney, Australia), Ken Koedinger (CMU, USA), Bruce McLaren (CMU, USA), Ma. Mercedes T. Rodrigo (Ateneo de Manila University, Philippines), Olga C. Santos (UNED, Spain), Eileen Scanlon (OU, UK), Mike Sharples (OU, UK), Erin Walker (University of Pittsburgh, USA), and Beverly Woolf (UMASS, USA).

In this paper, we first summarise the background (including the ethics of AI in general, and the ethics of educational data) and introduce the ethics of AI in education. Next, we outline the methods used in this project, before going on to summarise the survey responses (NB all responses quoted in this paper are verbatim, with any words that have been added or deleted for clarity marked by square parentheses or ellipses as appropriate). We finish with a discussion of the issues raised by the survey, and a suggestion for next steps.

Background

The AIED conferences in 2018 and 2019 hosted a workshop called *"Ethics of AIED. Who Cares?"* (Holmes et al. 2019a; Holmes et al. 2018). Although the workshop discussions were engaging and productive, the small number of attendees suggested a disappointingly low level of interest in the topic from the broader AIED community. While other AI communities are increasingly attending to ethical considerations around the design and deployment of AI-based technologies, ethical dimensions of AIED do not yet appear to be a central area of focus for many in the AIED community. Since Aiken and Epstein published their ethical guidelines two decades ago to begin a conversation around the ethics of AIED (Aiken and Epstein 2000), there has been a

striking paucity of published work in the AIED community that explicitly focuses on ethics. In any case, the potential impact of AIED designs and methods of deployment on students, teachers and wider society appears yet to be fully worked out.

Nonetheless, it is generally accepted within the community that AIED raises far-reaching questions with important ethical implications for students, educators, parents, policymakers, and other stakeholders. Ethical concerns permeate many of the community's core interests, including but not limited to: accuracy of diagnoses of learners interacting with AIED systems; choices of pedagogies employed by AIED systems; predictions of learning outcomes made by those systems; issues of fairness, accountability, and transparency; and questions related to the influence of AI and learning analytics on teachers' decision making.

In many ways, AIED is itself a response to each of those issues and more – with most being addressed, one way or another, in work conducted in the varied research subdomains of AIED. For example, questions around data ownership and control over their interpretations have long been recognised as critical in AIED (e.g., in the context of open learner modelling: Bull and Kay 2016; Conati et al. 2018). However, what is currently missing is a basis for the meaningful ethical reflection necessary for innovation in AIED, to help researchers determine how they might best contribute towards addressing those challenges explicitly. This requires a deep engagement both with recent ethics related debates and frameworks in other AI research subdomains, and with ethical questions specific to the AIED subdomain itself (e.g., issues around pedagogy), many of which remain unasked and unanswered (although there are some exceptions - e.g., Aiken and Epstein 2000; Friedman and Kahn 1992; Holmes et al. 2019b; Sharkey 2016).

The Ethics of AI in General

As with any transformative technology, some AI applications may raise new ethical and legal questions, for example related to liability or potentially biased decision-making. The ethics of artificial intelligence in general has received a great deal of attention, by researchers (e.g., Boddington 2017; Floridi 2019; Jobin et al. 2019; Whittaker et al. 2018; Winfield and Jirotka 2018) and more widely (e.g., the European Union 2019; the UK's House of Lords 2018; UNESCO 2019; and the World Economic Forum 2019), with numerous other AI ethics initiatives emerging in recent years (e.g., Ada Lovelace Institute 2019; AI Ethics Initiative 2017; AI Now Institute 2017; DeepMind Ethics & Society 2017; Future of Life Institute 2013; The Institute for Ethical AI & Machine Learning 2018).

All of these efforts principally focus on data (involving issues such as informed consent, data privacy, and biased data sets) and how that data is analysed (involving issues such as biased assumptions, transparency and statistical apophenia – finding patterns where no meaningful, causal patterns are present). The Montréal Declaration for Responsible Development of Artificial Intelligence (2018), for example, offers a comprehensive approach involving ten human-centred principles, encompassing: well-being, respect for autonomy, protection of privacy, solidarity, democratic participation, equity, diversity, prudence, responsibility, and sustainable development. No such declaration currently exists for the specific issues raised by AIED.

The Ethics of Educational Data

Similarly, the ethics of educational data and learning analytics has also been the focus of much research (e.g., Ferguson et al. 2016; Slade and Prinsloo 2013; Potgieter 2020). This work is extensive, far too wide to summarise here; however, some key issues can be noted. First, because the field is still emerging, exactly what an ethics of learning analytics should include remains the subject of debate: *'the ethical and privacy aspects of learning analytics are varied, and they shift as the use of data reveals information that could not be accessed in the past'* (Ferguson et al. 2016, p. 5). Second, the ethics of learning analytics involves several types of questions, including but not limited to: informed consent and privacy, the interpretation of data, the management of data, and perspectives on data (e.g., institutional versus individual); as well as on much broader issues such as power relations, surveillance, and the purpose of education (Slade and Prinsloo 2013). Third, it has been argued that *'educational data mining* [...] *is not the superconductor of truth that some of its proponents believe* [...] *and the transformative impact that it will have on the autonomy of learners is cause for concern'* (Potgieter 2020, pp. 3, 6).

The learning analytics community have endeavoured to agree principles against which learning analytics research and practice can judge itself and be judged. The DELICATE checklist, for example, comprises guidance centred on determining added value and the rights of participants, being open about intentions and objectives, legitimising the collection of data, involving all stakeholders including the data subjects, ensuring consent is genuinely informed and freely given, ensuring that data is truly anonymised, establishing and implementing procedures to guarantee individual privacy, and adopting clear and transparent obligations with any external agencies that might be involved with the data (Drachsler and Greller 2016). In fact, the clear overlaps between learning analytics and AIED, that are centred on educational data, suggest that an ethics of AIED might usefully draw on approaches such as the DELICATE checklist. However, there are also clear differences between the two fields, *"with an emphasis on agents and tutors for AIED,* [...] *and visualization for LA"* (Labarthe et al. 2018, p. 70). This active engagement/passive representation distinction, although radically oversimplified, suggests that a comprehensive ethics of AIED is likely to have additional requirements.

The Ethics of Other Related Research Areas

Ethics have also received attention in related research areas such as user modeling, e-learning environments, and intelligent agents. For example, in the user modeling community, issues centred on security and privacy (e.g., Schreck 2003), inspectability (e.g., Zapata-Rivera and Greer 2004), and privacy (e.g., Kobsa 2007) have long been considered. Most recently, dedicated workshops have been held at leading conferences (e.g., FairUMAP, Mobasher et al. 2020), with contributions exploring issues such as fairness (Sacharidis et al. 2020), transparency (Schelenz et al. 2020), and biases (Deshpande et al. 2020). Researchers in e-learning are also interested in ethical issues raised by the systems that they build; issues such as equity and diversity, surveillance and consent, identity and confidentiality (e.g., Anderson and Simpson 2007), and student privacy: *"when an observer* [which could be an automated system] *monitors*

someone's behavior with full knowledge of their identity, the person being monitored does not enjoy any privacy" (Anwar and Greer 2012, p. 63). However, it is notable that other researchers in e-learning appear more interested in the ethical practices of the students, such as cheating and fraud (e.g., Gearhart 2012). Finally, researchers and developers of intelligent agents such as chatbots are also increasingly focusing on the ethical issues raised by their work (e.g., Murtarelli et al. 2020; Richards and Dignum 2019). Most recently, the World Economic Forum has launched 'RESET', a framework and set of principles for governing responsible use of conversational AI (WEF 2020) – which, although focused on healthcare, have self-evident potential for the ethics of intelligent agents in education. The RESET principles focus on safety/non-maleficence, efficacy, data protection, human agency, accountability, transparency, fairness, explainability, integrity and inclusiveness.

The Ethics of AIED in Particular

As with AI in general, concerns exist about the large volumes of data collected to support AIED (such as the recording of student competencies, inferred emotional states, strategies and misconceptions). Who owns and who is able to access these data, what are the privacy concerns, and who should be considered responsible if something goes wrong?

Other major AIED ethical concerns, again as with AI in general, centre on the computational approaches. How should the data be analysed, interpreted, shared and acted upon? How should the biases (conscious or unconscious), that might impact negatively on the civil rights of individual students, be prevented or ameliorated – especially given that the scale of AIED in the coming years is likely to amplify any design biases (e.g., about gender, age, race, social status, income inequality…)? Finally, as the Facebook and Cambridge Analytica data scandal showed, data is vulnerable to hacking and manipulation: *'it's impossible to have personal privacy and control at scale, so it is critical that the uses to which data will be put are ethical – and that the ethical guidelines are clearly understood'* (Tarran 2018, pp. 4–5).

However, the ethics of AIED cannot be reduced to questions about data or computational approaches alone (Holmes et al. 2019b). AIED research also needs to account for the *ethics of education*, which, although the subject of decades of research, is all too often overlooked. For example, AIED research needs to address explicitly issues such as (1) the purpose of the learning (e.g., to prepare students to pass exams or to help them self-actualise), (2) the choice of pedagogy (with a common approach, instructionism, being contested by the learning sciences community), (3) the role of the technology with respect to teachers (to replace or augment human functions), and (4) access to education (often seen by the community through the ethical dimension of fairness and equity). In addition, there remains limited research into what teachers and students actually want from AIED systems - such as requirements around student agency and privacy about which teachers and students might not agree (Holstein et al. 2019). Furthermore, where AIED interventions target behavioural change (such as by 'nudging' individuals towards a particular course of action), the entire sequence of AIED enhanced pedagogical activity needs to be ethically warranted in the context of the broader activities within which AIED systems are being deployed.

To highlight just some of the potential breadth of issues, AIED ethical questions include:

- How does the transient nature of student goals, interests and emotions impact on the ethics of AIED?
- How can K12 students give genuinely informed consent for their involvement with AIED tools?
- What are the AIED ethical obligations of private organisations (developers of AIED products) and public authorities (schools and universities involved in AIED research)?
- How might schools, students and teachers opt out from, or challenge, how they are represented in large datasets?
- What are the ethical implications of not being able to easily interrogate how some AIED deep decisions (e.g., those using multi-level neural networks) are made?
- What are the ethical consequences of encouraging students to work independently with AI-supported software (rather than with teachers or in collaborative groups)?

Notably, as mentioned above, some guidelines were proposed almost 20 years ago (Aiken and Epstein 2000), but have not been widely adopted by the AIED community. Aiken and Epstein start with a negative and a positive meta-principle related to the impact of AIED on 'dimensions of human being'; with the dimensions ethical, aesthetic, social, intellectual, and physical, together with psychological traits such as 'the individual's ability to lead a happy and fulfilling life'. Aiken and Epstein's negative meta-principle is that '(AIED) *technology should not diminish the student along any of the fundamental dimensions of human being*.' The positive meta-principle is that '*AIED technology should augment the student along at least one of the fundamental dimensions of human being*.' They go on to provide and discuss ten 'fundamental principles' for educational technologies that incorporate AI methods. Some of these (such as 'avoid information overload') are essential tenets of good user experience and effective pedagogy for most educational technologies. The principles that are most specific to the application of AI in education are: '*7. Develop systems that give teachers new and creative roles that might not have been possible before the use of technology. Systems should not attempt to replace the teacher.*' and '*10. Avoid glorifying the use of computer systems, thereby diminishing the human role and the human potential for learning and growth.*'

Method

Broadly, the aim of this work is to galvanise the AIED research community's engagement with the ethics of its domain. Accordingly, it was first important to establish community members' existing beliefs and understandings of the ethics of AI applied in educational contexts, to which end a survey of AIED researchers was developed and conducted. The survey questionnaire comprised 10 substantive open questions (Appendix 1), which covered the following themes: (i) what it means for AIED researchers to consider the ethics of their work, (ii) pertinent AI and education ethical issues, and (iii) what distinguishes the ethics of AIED from the ethics of other AI domains. The

research was approved by the Open University's Human Research Ethics Committee (HREC/3339/Holmes: 02/08/2019).

Potential invitees for the survey were identified by reviewing contributions to the *International Journal of Artificial Intelligence in Education* and the proceedings of the *International Conference on Artificial Intelligence in Education*, both from the last 10 years. The names of researchers with the most citations, together with those researchers otherwise known by reputation (e.g., they hold or have held a central role in the *International Artificial Intelligence in Education Society*), were collated. To this list were added some mid- and early-career researchers who had already made a noteworthy impact in the community (e.g., they had received a Best Paper award at an AIED conference). In total, the list comprised 60 candidate survey invitees, all of whom were invited to participate in the survey. An invitation email and up to three reminder emails were sent to each invitee, 17 of whom completed the questionnaire.

At the beginning of the questionnaire, respondents were asked to give explicit permission for their names to be given in the proposed paper, and for any of their responses included in the paper to be attributed to them (this approach draws on Guideline 41 of the British Educational Research Association's Ethical Guidelines for Educational Research 2018). This was both to ensure transparency and to maximise impact on early- and mid-career AIED researchers. All respondents gave their permission. In addition, before the paper was submitted for publication, the respondents were given the opportunity to check, amend, agree, and/or delete any quote that was attributed to them. No respondent requested any substantive change. Finally, to further acknowledge their contribution, the respondents were given the opportunity to be named as a co-author of the paper. Six respondents took this up (Holstein and Buckingham Shum also made substantial contributions to the writing of the paper).

Results

Paying Attention to Ethics in AIED

Survey respondents were first asked to reflect upon whether and why they agree or disagree with the statement, *"AIED researchers do not pay sufficient attention to the ethics of AIED."* All 17 respondents answered this question, revealing overall agreement that there is room for improvement. For instance, Buckingham Shum noted that while *"there are vibrant communities and journals"* focusing on the ethics of technology, these exist as *"a 'parallel universe' separate from the AIED circuit."* Similarly, Holstein observed that when the topic of AIED comes up in communities other than AIED itself, *"ethical considerations tend to be the first topics raised by those outside the community,"* but that *"within AIED* [...] *this is not so."* Several respondents noted that the field lacks shared frameworks for thinking through ethical issues in AIED. As Walker noted, *"collectively we lack the training in frameworks and principles that would allow us to grapple with major ethical questions of AIED research."*

However, five out of 17 respondents reported that they perceive growing urgency and attention to ethical issues within the AIED research community. For example, Scanlon noted that *"there is much discussion now of ethics because of high profile examples in AI which are hitting the media* [...] *ethical consideration*[s] *are becoming*

not just part of the academic community's discourse but part of wider public discussion." Similarly, Bittencourt suggested that *"since* [there is so much] *news and arguments about the implications of AI, researchers are starting to think about ethics in AIED."* Buckingham Shum framed growing concerns around the ethics of AIED as a consequence of the field's growth in real-world applications: *"It is only relatively recently with 'adaptive products' hitting the mainstream that AIED has found itself being touched by the broader public spotlight. Until then, it was an academic community, with such a small user base that nobody was asking ethics questions."*

Only one participant out of 17 outright disagreed that AIED researchers pay insufficient attention to ethics. However, this participant's response focused exclusively on AIED *research ethics*, noting that institutional ethics boards in university settings ensure that researchers consider certain ethical dimensions of their research practice. Four of 17 respondents described themselves as *"on the fence,"* or stated that they agreed only *"to an extent."* Koedinger noted that *"many of us got into AIED because we wanted to improve education, and especially for students who most need the help of free public education,"* characterizing the pursuit of this goal as *"doing ethics, which goes beyond research about* [ethics]*."* Aleven similarly said that *"one suspects that many AIED researchers are fundamentally motivated by ethical concerns,"* while acknowledging that *"so far the field has not often addressed* [ethics] *explicitly or discussed it publicly."* Other respondents pointed out that *"this is a problem inherent to many fields of technology research and not solely AIED"* (Walker). In line with this view, Sharples characterized the practice of AIED as being *"primarily led by technology,"* noting that *"its impacts on stakeholders, including direct users, those indirectly affected and society at large has been largely ignored."* Finally, some respondents commented that this is a complex question to answer. McLaren noted that it may be difficult to make generalizations across the whole of the AIED community, as *"some researchers are very cognizant and sensitive to ethical issues* [...] *others, not as much."* Other respondents emphasised that the AIED community has attended to certain aspects of ethics (e.g., data privacy and research ethics), although historically it has attended less to others (Rodrigo).

In a follow-up question, survey respondents were asked to reflect upon *why* it is important for AIED researchers to pay careful consideration to the ethics of their work. Respondents highlighted the role AIED research may play in shaping what education looks like for the next generation of students. For example, Cukurova emphasised the potential for *"long term adoption of technology."* McLaren noted that AIED research *"could have an impact on students, their education, and thus the future of mankind."* Walker, Rodrigo, and Holstein emphasised the moral responsibilities that come with AIED being a design discipline, not solely an empirical science. As Walker put it, *"we are shaping a future where technology can personalize instruction to learners' unique needs, and should be more involved in ensuring that future is a positive one."* Similarly, Rodrigo highlighted that AIED researchers are engaged in *"trying to create systems that can assist in the formation of the next generation. These students have to be treated fairly, with respect, with consideration, and their needs as individuals have to be attended to, to the extent possible."*

A number of respondents emphasised that AIED researchers have a responsibility to make their *"research more accessible, and more transparent"* (Buckingham Shum), and to *"be particularly mindful of the ways that people interpret our results"* (Kay). As

Gobert noted, *"generalizing results beyond the demographics for whom the study was designed could mean that students might get interventions that are not suitable for them."* du Boulay distinguished between *"two main areas of concern"* for AIED research: *"the danger of harm to participants,"* including (1) *"harm arising from decisions that the AIED system might make,"* and (2) *"the careful curation of data derived from the research,"* including the related issue of *"the models of individuals that the AIED system might build."* Other responses focused specifically on the unique sensitivity of the data collected and used by AIED systems. For example, Boticario said *"Data are no longer data because there is so much information available on anyone. Data are people themselves and should be treated as such* [...] *we are* [providing] *personalised support to our students based on* [an] *increasing body of data, which is starting to include personal and affective information. This information is by its nature sensitive and demands a careful consideration and management."* Finally, one respondent noted that the need for AIED researchers to attend to the ethics of their work had been *"argued in a journal article* [in] *IJAIED back in 2000,"* (Aiken and Epstein 2000) and asked, *"how has it changed?"*

Most Important Ethical Issues for AIED

When asked '*What are the most important ethical issues for AIED?*', nine respondents directly referenced 'data', with several more identifying issues overlapping with the governance of data. Many of the issues cited are not specific to education settings. These included: (1) Data ownership and control, both within and outside the context of research: "*In a research context, the data normally "belongs" to the researcher. Once a system is deployed in normal teaching, the issue of the ownership of this data is much more problematic*" (du Boulay). (2) Expectations of privacy, while recognising that "*privacy is more than a binary value, it has many nuances and shades*" (Boticario). (3) Limitations of data, bias and representation, "*where information is inferred without noticing the existing limitations or the actual evidence behind results*" (Boticario). (4) Consequently, "*the management of the data collection, where users need to be properly informed*" (Santos). (5) The transparency and intelligibility of decisions: "*user agency, and user awareness of how the technology functions are important issues"* (Walker). Emerging technologies can bring new dimensions to these issues, with one respondent asking "*How to treat facial recognition technology?"* (Scanlon).

In addition, several respondents identified properties of AI in *education* which, in combination together, make education distinct to other sectors: "*All the generic concerns around big data and algorithms apply to education. What's more interesting are the features that make the formal education system, and learning as a phenomenon, distinctive from the other sectors of society being impacted by data science"* (Buckingham Shum). These issues relate not just to the ethics of the development of AIED tools in and of themselves, but to their educational context – when, how, and to what end, an AIED tool is being used.

Unsurprisingly, the implications of AIED for teaching and learning was central to many of the respondents' concerns. The "*quality of education*" (Aleven) enabled by AIED tools was highlighted. There was a recognition of the pedagogical choices inherent within the design of AIED tools, with one respondent emphasising a desire for "*designing a pedagogy that empowers learners and teachers*" (Sharples), while

another arguing that the scope of many AIED tools remains narrow and therefore there may be elements of education that are better off without AIED, asking: "*What do we risk losing through automation, and are we prepared to preserve those aspects that should be preserved (e.g., intentionally 'not' automated)? Are we diminishing the quality of education, overall, by conveying that we have reached human parity in AI-based teaching systems (but failing to emphasise that we have done so on a very narrow set of teaching tasks, whereas good education involves much more than this)?*" (Holstein).

A further set of concerns for several respondents centered on the experience of users. These included reflections on agency, transparency and intelligibility in education contexts. For example, "*making the teaching and learning actions transparent to learners and teachers*" (Sharples), or "*avoiding alienation on the part of teachers, students, and other stakeholders*" (Aleven). These also included reference to equity and fairness. For example, "s*ome students may need to work longer than others; ideally, every student should reach the same level of mastery of the content they need to learn*" (Woolf), or "*are we diminishing the quality of education for 'certain groups of students', in cases where 'business as usual' would have worked better for these groups?"* (Holstein). That respondent also identified a risk at a societal level that users of AIED could be subjected to "*cultural imperialism as we scale up AIED systems to a diverse range of contexts*" (Holstein).

A final group of concerns related to the ethics of AIED *research*, rather than its applications and uses. These covered the need to treat research subjects – particularly children – ethically and fairly (McLaren), reporting results while respecting confidentiality and anonymity, particularly given potential to re-identify data (Santos, Boticario), and data collection (Gobert). In addition, the potential for AIED research to be (mis)used by companies was identified: '*How will our outputs be used (or misused) by more profit-driven entities? How do we manage misuses of our outputs?*' (Holstein).

Most Ignored Yet Critical Ethical Issues for AIED

When asked what were the most ignored ethical issues for AIED, several of the respondents (Bittencourt, Boticario, du Boulay, Santos, Scanlon, Sharples) again emphasised issues centred on data, for example: (1) The problems of data collection: *"What I think it is* […] *critical is how to approach the need to collect data in a real world scenario when you need to compare proposals and you think one of them might be bad for the students, or give them less opportunities for learning"* (Santos). (2) *"The lack of clarity of what happens to the data collected by AIED systems"* (du Boulay). (3) Data ownership and control: *"The student is providing data for their benefit, and thereof they should have to control what data and what usage of that data is to be allowed. No one but them must have the control or right to manipulate those data for anything else but to support their learning"* (Boticario). (4) Data anonymity: *"In many journal papers, the reporting of the data is anonymized, yet participants are able to recognize themselves in the reporting because there is only one woman who is 26 years old and has 3 years of experience as* [a] *teacher"* (Santos). In a similar vein, other researchers mentioned issues around transparency: *"How to make the teaching decisions and the system's model of the learner visible and inspectable"* (Sharples).

Many respondents (including Aleven, Holstein, Koedinger, Rodrigo, Walker and Woolf) raised issues around representativeness and equity: are we diminishing quality education for certain groups, are AIED systems biased against some groups, and are some other groups being ignored? The many complementary comments are worth quoting: *"I think because it is so difficult to collect representative data sets and because funding is structured so as not to reach out to a multitude of populations, it is accepted that data sets will always be limited"* (Rodrigo). *"The challenge of designing equitable AIED technologies that do not show bias against a particular group is not discussed much in the community, and probably should be engaged with"* (Walker). *"Are we diminishing the quality of education for 'certain groups of students', in cases where 'business as usual' would have worked better for these groups?"* (Holstein). *"The degree of personalisation for some students might be less than that for others, which could lead to inequitable outcomes. Personalised instructional strategies could ignore under-served/ underrepresented/minority students and fall short of providing fair learning opportunities to each student"* (Woolf). *"It seems inevitable that giving the same instruction to students who are well prepared for that instruction versus students who are not well prepared for it will lead to different results.* [...] *Should we now conclude that this instructional intervention is unethical, as it seems to be increasing inequality?"* (Aleven). *"There could be more projects that have direct impact on disadvantaged learners"* (Koedinger).

Finally, issues were raised over AIED systems and human agency, for example *"How to embed pedagogies of learner and teacher empowerment into AIED systems"* (Sharples). Areas that need specific attention are: *"The ongoing misplaced belief that AIED systems can wholly replace human teachers and still enable effective and socially responsible learning."* (du Boulay), *"The lack of human agency in fully autonomous AI in Education systems"* (Cukurova), and *"What* […] *we risk losing through automation,* […] *are we prepared to preserve those aspects that should be preserved (e.g., intentionally 'not' automated)?"* (Holstein).

What Distinguishes the Ethics of AIED from Other AI Applications?

Survey respondents were asked to reflect on what distinguishes the ethics of AIED from other AI applications/domains and to identify the most important 'educational' ethical issues for AIED. The majority of responses identified how issues common to the 'ethics of AI' manifest in educational contexts - particularly issues of consent, transparency, control over data and algorithmic bias. In particular, the application of AI to children and young people was noted as a distinguishing feature of AIED (e.g., McLaren). This was not only associated with issues of consent, but also the longer-term individual and social impacts of engaging with young people during a formative period: *"We are influencing impressionable young people, and that comes with a moral obligation to be as correct and proper as possible"* (Rodrigo). Despite a general recognition of the potential long-term impacts of AIED, two respondents (du Boulay, Santos) also noted that AIED systems do not make the same 'life and death' decisions typical of other AI applications, such as autonomous vehicles or medicine: *"a bad treatment in learning might imply more effort from the student or more time, but it does not seem as critical as a bad treatment in medicine or a bad decision in an autonomous car, which can cause death"* (Santos). This

reveals a tension between how researchers frame the ethical issues and potential 'harms' associated with AIED.

Another issue noted by the respondents was the value of transparency in AIED systems, with one respondent suggesting that teachers might prefer an AI system that is less accurate but offers more transparency around decision-making "*over an AI system that is perfectly accurate but not transparent"* (Cukurova). Others referred to the importance of giving learners agency and control over their data, and the potential impact of algorithmic bias on different sub-populations and demographics of learners. There was particular concern that biased data sets have the potential to benefit more advantaged groups of learners: *"Instructional strategies that aim to benefit all learners might disproportionately benefit more advantaged groups of learners. Our community needs to learn to integrate fairness-promoting algorithms into the system and deploy them for real students"* (Woolf).

Markedly less common were responses that identified issues specific to the ethics *of* education, and how these intersect with the ethics of AI more generally. Choice of pedagogy was mentioned by a few respondents, who acknowledged that AIED embodies assumptions about which instructional strategies are 'best' for learners. For example,*"pedagogy driven AI in Education is often considered as good regardless of* [whether] *the pedagogy chosen to drive AI is an appropriate one or not in the first place"* (Cukurova). One respondent also referred to issues relating to the accuracy and validity of learning assessments more generally: *"classifying students in terms of educational tests has to consider the inherent ambiguity and variability in the measure, but for computer scientists we usually consider them as rigid labels"* (Santos). Brief reference was also made to the particular responsibilities of educators and the dependent relationship between learners and educators/teaching systems.

Advice for Teachers, Faculty Members and Professors about Teaching the Ethics of AIED

When asked for their advice on how best to teach AIED students to properly address the ethics of AIED, some respondents acknowledged that they did not feel sufficiently qualified: *"To be quite honest, I do not know.... I would like to note that very few of us are trained in ethics!"* (Aleven). *"I feel like I need to better educate myself about this (and in fact, I joined a committee to introduce an ethics curriculum in my department in part to try to better inform myself)"* (Walker). One suggestion for how this might be addressed was to: *"bring together voices from academics who study ethics, those who build systems and those who will use them."* (Scanlon). Aleven broadly agreed: "*To properly address ethical issues, we probably need to look at a bigger picture than we are used to doing as AIED researchers. And* [...]*, we should avoid armchair ethics and work with experts in ethics"*. Nonetheless, a starting point might be for academics to 'practise what they preach': *"Ethics need to be modeled in order to be taught. We academics have to conduct ourselves and our research practices as ethically as we know how"* (Rodrigo).

Other respondents suggested that academics "*should start with the philosophy of technology use and adoption in general"* (Cukurova), and should *"learn from the history of AIED, to understand how AIED implements an explicit or implicit theory of teaching and learning, to understand how that affects the learning outcomes and life*

opportunities of learners, and to explore how this relates to [the] *ethics of research in education"* (Sharples). Starting with the wider context and previous literature is important perhaps because *"we need better guidance than we currently have. A checklist, with items linked to the relevant foundation documents"* (Kay). *"Maybe a good starting point is to collect some interesting examples of how research can help with ethical issues in AIED"* (Aleven).

A common thought was that students would learn best through direct experience: *"Don't just 'build ethics into the curriculum'. Have students learn by doing. Make ethical considerations a core part of the evaluation criteria (e.g., for student deliverables in project-based courses). Make class discussions around ethics part of the actual formative evaluation / feedback process for student projects"* (Holstein). *"I would also suggest that they should give actual hands-on experience with AI systems, let their students build, test, play with AI tools in controlled conditions to understand potential ethical issues through experience"* (Cukurova). *"Introduce them to accessible frameworks (such as those beginning to emerge in AI FATE* [e.g., https://facctconference.org] *work) and develop exercises that help people to apply them. This will in turn refine them"* (Buckingham Shum) .

Other respondents took this approach an intriguing step further: students might benefit from experience of AIED tools that have been developed without addressing any ethics: *"I think it could be an interesting approach to ask them* [to] *be part of a study where ethics are not considered, so they can perceive the need to take care of it"* (Santos). *"Mak*[e] *them investigate a scenario where those principles are not followed and have them research all possible consequences. This way they could figure out not just the basic principles involved but also possible flaws or missing considerations"* (Boticario).

Advice for Early- and Mid-career Researchers About the Ethics of AIED

Survey respondents were asked what advice they would give to early- and mid-career researchers about the ethics of AIED. The 16 responses to this question encouraged early- to mid-career researchers to take ethics seriously (Bittencourt, Boticario, du Boulay, Gobert, Holstein, Koedinger, Rodrigo, Sharples, Woolf), to actively participate in discussions about what ethical practice in AIED should look like (Aleven, Kay, Scanlon, Sharples), to critically reflect upon and question their own research practices (Boticario, Cukurova, Holstein, Walker), and to engage with relevant expertise outside of the AIED community (Buckingham Shum, Cukurova, Santos).

Sharples encouraged early- to mid-career researchers to actively contribute to the *"development of a code of ethical practice in AIED,"* and to abide by this code. A number of respondents urged researchers towards more critical reflection around their own research practice. Walker advised researchers to *"think about the implications of what you're doing and be able to critically engage with them."* Similarly, Boticario said, *"we have the *responsibility* to keep our eyes open and respond to what is happening in the world around us. As AIED researchers we are constantly changing the trajectories of others' lives.* [...] *If we cannot take responsibility for our actions, we do not deserve to remain in this field."* Holstein advised that *"we are a generation* [of researchers] *that cannot afford to ignore the ethics of AIED,"* noting that *"generations that are trying to push an innovative technology into the world have the luxury of*

thinking primarily about hypotheticals," whereas *"the generation that carries forward this work once the systems are already being* [used] *and impacting lives at scale"* must reckon with more immediate ethical concerns.

Finally, several respondents urged early- to mid-career AIED researchers to expand beyond the AIED literature, and to engage with the broader research community outside of AIED. For example, Buckingham Shum encouraged AIED researchers to *"consider engaging with academics in critical data studies and tech-ethics."* Similarly, Cukurova encouraged researchers to *"read fundamental philosophy as most ethical questions come down to philosophical discussions and frameworks that are driven by certain values."*

Should the IJAIED Require an Ethics Statement for all Articles?

Respondents were asked whether they agreed with the suggestion that all articles published in this journal should be accompanied by a statement explaining how the authors have fully accounted for the ethics of their AIED work. Of the 16 respondents to this question, only two disagreed, but mainly because *"the ethics of AIED are not yet defined sufficiently to make this useful or viable"* (du Boulay). This was echoed by many of the respondents who did agree (Bittencourt, Boticario, Buckingham Shum, Gobert, Kay, Sharples, Walker), who for example pointed out that *"we need much more nuanced conceptions of what 'ethical' means"* (Buckingham Shum), and that the journal would first need *"to indicate clearly what is meant by 'fully accounted for the ethics'"* (Sharples).

Nonetheless, the respondents were broadly in favour. Santos, for example, noted *"I think it would help to get more awareness and try to do things right. Maybe it can be asked to have a short section at the end of the paper that the authors explain in their own words how they have covered the ethics in the research reported. It would be something similar to the bios some journals have, or the highlights section*." Other tangible suggestions included: *"Perhaps the Journal can ask authors to upload a copy of their ethics clearance"* (Rodrigo), and *"Authors should be required to provide statements of both positive and negative broader impacts, along with a discussion of trade-offs. Reviewers should then be specifically asked to evaluate these statements, and should be incentivized not to let shallow, 'ticking the box'-style slide."* (Holstein).

Woolf brought the discussion back to AIED specifics: *"Researchers should determine whether the current adaptive tutor is biased in every applied paper published. For* [example,] *examine the algorithm or the model or knowledge tracing used to track student progress. Does an algorithm select problems that allow students to demonstrate skill mastery before moving on to new content? All authors should evaluate whether their teaching system learned on training data that introduced bias along certain socio-demographic categories."*

Another challenge noted by respondents was how to ensure that such a policy was meaningful: *"I do think this is a good idea. As to how it won't become a tick box, I am totally unsure"* (Gobert). In any case, Walker noted, *"in practice I do think it is difficult to enforce"*. Others proposed ways in which the spirit of the statement might more effectively be addressed: *"My suggestion is to change that statement into a clear and easily affordable questionnaire, where the most basic issues can be specified"* (Boticario); *"Publishing a more specific list of ethical principles you expect people*

to engage with and affirm may be an intermediate step" (Walker); *"I think it would be very useful to create such a check list of issues to consider"* (Kay).

Discussion

The survey results presented in this paper indicate that AIED researchers recognise the importance and value of engaging with the ethics of their work (e.g., *"We need to be sure that our work is of both the highest standards of research and the highest standards of ethics."* McLaren). Nonetheless, there are nuances of opinion with respect to what this might include and how it might be best achieved. With this in mind, in this section, we first summarise key emergent themes and identify some issues that we believe are missing from the discussion; then we propose a draft of an initial AIED ethics framework to help galvanise further debate.

As noted previously, there is no doubt that all respondents do wish to explore, better understand and engage with the ethics of the design and application of AI in educational contexts. However, some respondents appear to believe that we, as a community, are already 'doing ethics' by virtue of operating with best intentions in the educational domain, which is in and of itself ethical (e.g., *"Many of us got into AIED because we wanted to improve education* [...]. *Pursuing that goal is doing ethics* [...]. *Yet, definitely, we could do more*." Koedinger). However, the reality is that *"no ethical oversight is required to deploy an elearning system or an AIED system as part of the normal teaching process"* (du Boulay).

Clearly, good intentions are not by themselves sufficient. As was acknowledged by several respondents, we need to understand more precisely the ethical risks and to be always on the lookout for unintended consequences that relate specifically to the pedagogical designs (including their readiness for real-world use) that are encapsulated in the AI systems we develop and deploy (e.g., *"One could envisage a time when judgements about progress through educational stages might be taken by AIED systems and these might be seriously damaging to individuals' progress and well-being if they were wrong."* du Boulay).

The precision of such an understanding needs to be expressed in *an actionable code of best practice* that the community can rely on in designing and deploying AIED technologies in diverse educational contexts. However, although there appears to be a clear appetite for some kind of ethics of AIED framework that would build upon university research ethics, the community also recognises that such a framework also needs to be distinct from the generally established research ethics approvals and procedures. In particular, such a framework would need to incorporate guidance addressing the many issues raised by respondents (including fairness, accountability, transparency, bias, autonomy, agency, and inclusion) specifically at the intersection of AI and the learning sciences, and ensure that AIED is ethical by design and not just by intention.

In this context, the respondents were also cognizant of the problems that such a framework might itself inadvertently entail. In particular, in line with Bietti (2020), having such a framework might both stifle innovation and lead to accusations of, or actual, in-house *"ethics washing"*. So, while the community recognises the critical importance of AIED's specific code of ethical practice, what is meant by ethical AIED

remains an open question (e.g., *"There is no agreement on the guiding principles."* Boticario). Thus, establishing a useful framework for the ethics of AI in education, although desirable and potentially useful, is likely to be a challenging and long-term task. Specifically, given the diversity of interpretations within the community, as indicated in the survey, it is likely that as well as incorporating presently known ethics-related issues and challenges, any ethics framework developed for the AIED context will need to allow for a degree of flexibility to incorporate new knowledge, new understandings and new ways of supporting learning and teaching, as our science, socio-cultural norms, values, and educational systems change over time (e.g., Porayska-Pomsta and Rajendran 2019).

A related important point that emerged from the survey is the observation that AIED is not merely an empirical science but also a design science (Holstein). We understand this to mean that AIED operates in both the theoretical and practical spheres of AI and the learning sciences, and further that as a design science it aims to create educational tools and interventions that are adapted to the demands of different specific users and contexts of use. Highlighting the practical and human-centric orientation of AIED brings into focus the fact that AIED is explicitly concerned with designing interventions which by definition aim to foster *positive behaviour change* in the users. This is especially important given that many of the intended users of AIED are at pivotal points in their intellectual development.

Adopting a design science orientation also points to another practical outworking of ethical principles – namely, the use of *human-centred design* methods that give stakeholders genuine agency in shaping digital tools, thus increasing the chances of producing tools that are usable, effective, organisationally acceptable, and ethically sound. However, the AIED community also has to be aware that, despite the best of intentions, human-centred design may itself, in some senses, be sometimes harmful to some users (Norman 2019). Nonetheless, a community is emerging around the critical adoption of human-centred design methods and tools for educational technologies powered by analytics and AI (e.g., Buckingham Shum et al. 2019). In common with the survey reported in this paper, they ask what makes the specific stakeholders and contexts of teaching and learning distinctive from other domains, that raise specific challenges for human-centred design.

All of these considerations further highlight related ethical implications for AIED. For example, *"The system forms an explicit or implicit model of the learner's knowledge and skills that it uses to make teaching interventions. That model may embody assumptions by the AIED designer as to appropriate methods of teaching and the abilities of the learner."* (Sharples). Thus, AIED's explicit aim to foster behaviour change in its users, adds additional pressure on the community, to embrace ethics and the related dimensions both as a necessity and as a moral obligation for the community. The flip side of this is that the field (alongside related fields such educational data mining, learning analytics, and user modelling) has a real potential to influence the educational systems at the frontline, precisely because it designs for and deploys in real-world contexts, and to contribute to broader approaches of designing for and deploying AI in other than educational human contexts.

In particular, it might be useful to consider how AIED research may contribute to broader debates about the ways in which AI might impact human cognition, decision-making and functioning. Given AIED's focus on human learning and development, it is

at least worth considering its potential role in informing those broader debates from its unique perspective of designing AI specifically for the purpose of influencing human cognition. AIED's ambition to support human learning and the field's proximity to the learning sciences, educational neuroscience and educational practice, likely affords a very human-centric, human-developmental understanding of ethics and related dimensions of fairness, transparency, accountability etc., than is afforded by the more general socio-political and legal considerations in the context of other AI subfields.

Furthermore, as well as contributing to the broader ethics of AI debates, there is a more specific question related to whether or not AIED as community might serve as a guiding discipline with respect to ethics best practices in non-academic areas such as the educational technology industry – something that Blikstein touched upon during his AIED 2018 conference keynote speech. Deciding how this might work best is non-trivial. To begin with, it might be worth considering a distinction between *"the ethics of AIED research"* (which should be informed by standards such as APA 2014 and BERA 2004), *"the ethics of AIED tools and practices in and outside of classrooms"* (for which, as we have noted, little has been published), and *"the ethics of commercial AIED products"* (for which nothing has so far been seen). For each of these, the community might benefit from drawing on the experience of other communities in related areas. In web accessibility research, for example, researchers have critiqued approaches that appear reasonable but that embody counterproductive assumptions that need to be problematised (e.g., Lewthwaite 2014).

The various stakeholders (developers, educators, policymakers) all need to be provided key information about the pros and cons of specific AIED technologies – perhaps something in the style of the list of ingredients and allergy warnings on food, or side-effects on medicines. Such key information could include both the known limitations (e.g. in terms of pedagogies, biases of interpretation, privacy etc.), as well as benefits that are likely to emerge from the use of specific AIED systems. Can such an open approach serve to better inform the users' choices and transparency of what we create? Would it allow the community to become more accountable, more in touch with the broader developments and debates in AI as well as educational practice? The idea is certainly worth exploring as it carries a potential for both making AIED more influential with respect to AI in Education policy and practice, and for informing AIED's future innovations and its standing in the broader scientific, educational practice and AI contexts.

First Steps towards a Framework

As noted above, the ethics of AI raises a variety of complex issues centred on data (e.g., consent and data privacy) and how that data is analysed (e.g., transparency and trust). However, it is also clear that the ethics of AIED cannot be reduced to questions about data and computational approaches alone. In other words, investigating the ethics of AIED data and computations is *necessary* but not *sufficient*. Given that, by definition, AIED is the application of AI techniques and processes in education, the ethics of AIED also as noted earlier needs to account for the ethics of education (Holmes et al. 2019b). Yet, while the ethics of education has been the focus of debate and research for more than 2000 years (e.g., Aristotle 2009; Macfarlane 2003; Peters 1970), it is mostly unacknowledged and unaccounted for by the wider AIED community.

Because of the rich history, there is inevitably insufficient space here to provide a comprehensive account of the ethics of education. Instead, we will simply identify some pertinent issues, each of which continues to be the subject of debate: the ethics of teacher expectations, of resource allocations (including teacher expertise), of gender and ethnic biases, of behaviour and discipline, of the accuracy and validity of assessments, of what constitutes useful knowledge, of teacher roles, of power relations between teachers and their students, and of particular approaches to pedagogy (teaching and learning, such as instructionism and constructivism).

The three foci identified – the ethics of data, computational approaches and education – together constitute the foundational level for a hypothesised comprehensive ethics of AIED framework (see Fig. 1). There is, however, a second level, which is concerned with the overlaps between the three foci (as illustrated in Fig. 1): the ethics of data in AI in general (which is, as discussed, an ongoing focus of much research; e.g., Boddington 2017), the ethics of data in education (also an ongoing focus; e.g., Ferguson et al. 2016), and the ethics of algorithms applied in educational contexts. This third overlap remains the least developed area of research.

However, a serious effort to develop a full ethics of AIED cannot be limited even to these six areas (data, computational approaches, education, and the overlaps between them). These constitute the 'known unknowns' of AIED ethics, but what about the 'unknown unknowns', the ethical issues raised by AIED that have yet to be even identified (i.e., those issues at the central intersection of data, computation and education, and the specific interaction between AI systems use and human cognition at the individual level, indicated by the question mark in Fig. 1)? Any sufficient ethics of AIED needs to involve horizon scanning, interdisciplinary conversations, explicitly taking into account insights from the

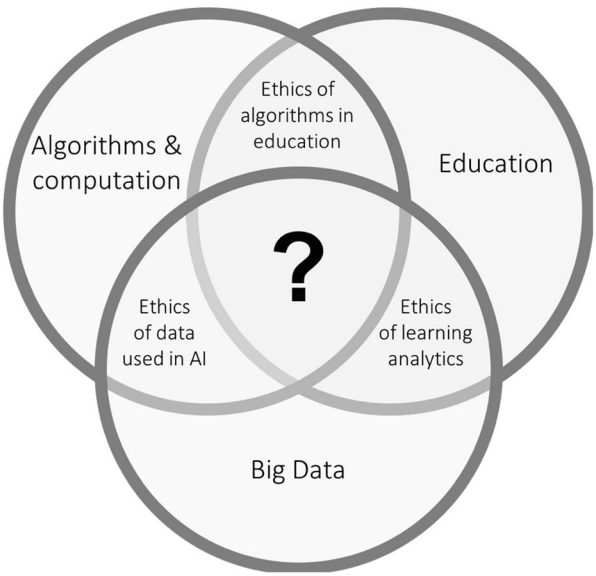

Fig. 1 A 'strawman' draft framework for the ethics of AIED

learning sciences, cognitive and educational neuroscience, and philosophical introspection. All of these are necessary to help us identify and explore the unknown unknowns, in order to establish a comprehensive framework for the ethics of AIED. In fact, establishing such a framework is only the first step in the process. If our efforts in this area are to have genuine and future use value for the AIED community, teachers, students, policymakers and other stakeholders, considerable effort then needs to be focused on *how* that framework might be best implemented in practice.

Finally, but no less importantly, we should recognise another perspective on AIED ethical questions: ethics is not just about stopping 'unethical' activities. Instead, the ethical cost of inaction and failure to innovate must be balanced against the potential for AIED innovation to result in real benefits for learners, educators, educational institutions and broader society. In other words, the ethics of AIED cannot just be preventative – it cannot just be about stopping researchers and developers from 'doing harm'. Instead, it needs to provide a proactive set of foundational guidance, within which to ground AIED research and development, which is both protective and facilitative, to help ensure the best outcomes for all stakeholders from the inevitable push of AI into educational contexts.

Conclusion

While the low numbers of researchers attending the Ethics of AIED workshops and responding to the survey reported in this paper suggest that the AIED community have a low level of interest in the ethics of AIED, the responses to this survey and the other papers in this SI indicate otherwise. Clearly, many AIED researchers do recognise the importance and value of engaging with the ethics of their work (indeed, there is no evidence of AIED work that is deliberately unethical). However, as the responses reported here have demonstrated, this engagement now needs to be surfaced, the nuances of opinion need to be discussed in depth, and issues around data, human cognition, and choices of pedagogy need to be investigated, challenged and resolved. In particular, the AIED community needs to debate the value and usefulness of developing an ethical framework and practical guidelines, to inform our ongoing research, and to ensure that the AIED tools that we develop and the approaches that we take are, in the widest sense, ethical by design. It is also clear that without a more targeted approach to the ethics of AIED, the work conducted by the community may remain largely invisible to the rest of the AI subfields and related policies, also potentially stifling the impact of the AIED research on the increasingly human-oriented, real-world applications of AI. With its deep understanding of the human users of AI and the AI's potential to support human learning and behaviour change, AIED offers critical perspective on the way that people interact with and change due to the interaction with AI systems, and on the potential benefits and pitfalls of such an engagement. The time is ripe to bring this perspective into the open and to allow for cross-fertilisation of AI science and approaches with the benefits for human learning and development as investigated for decades within the AIED field as the guiding principles for AIED and beyond.

APPENDIX 1: Survey questions

1. "AIED researchers do not pay sufficient attention to the ethics of AIED." Explain why you agree or disagree with this statement. Where appropriate, please give illustrative examples.
2. Why is it important for AIED researchers to pay careful consideration to the ethics of their work?
3. What are the most important ethical issues for AIED?
4. What are most often ignored yet critical ethical issues for AIED?
5. What distinguishes the ethics of AIED from the ethics of other AI applications/domains?
6. The ethics of AIED is often reduced to the ethics of AI (mainly the ethics of data and algorithms), while the ethics of education are often ignored. What are the most important education ethical issues for AIED?
7. "All articles published in The International Journal of Artificial Intelligence in Education (IJAIED) should be accompanied by a statement explaining how the authors have fully accounted for the ethics of their AIED work." Please explain why you agree or disagree with this statement. If you agree with the statement, please explain how the journal can ensure that this requirement does not become a worthless tick-box exercise.
8. What advice would you give to teachers/faculty members/professors about how best to teach/encourage their AIED students to properly address the ethics of AIED?
9. What advice would you give to early- and mid-career researchers about the ethics of AIED?
10. Please use the following box to give any other comments that you would like to make about the ethics of AIED.

Acknowledgements The authors would like to gratefully acknowledge all of the respondents who freely gave their time and expertise to complete the survey.
Code Availability N/A

Funding No funding was required or received for this research (i.e., the authors were funded as part of their standard affiliation duties).

Data Availability The full survey responses will be made available on request. The respondents all gave consent for this to happen.

Declarations

Conflict of Interest Seven of the authors were survey respondents.

References

Ada Lovelace Institute. (2019). https://www.adalovelaceinstitute.org

AI Ethics Initiative. (2017). https://aiethicsinitiative.org

AI Now Institute. (2017). https://ainowinstitute.org

Aiken, R. M., & Epstein, R. G. (2000). Ethical guidelines for AI in education: Starting a conversation. *International Journal of Artificial Intelligence in Education, 11*, 163–176.

Anderson, B., & Simpson, M. (2007). Ethical issues in online education. *Open Learning: The Journal of Open, Distance and e-Learning, 22*(2), 129–138. https://doi.org/10.1080/02680510701306673.

Anwar, M., & Greer, J. (2012). Facilitating trust in privacy-preserving E-learning environments. *IEEE Transactions on Learning Technologies, 5*(1), 62–73. https://doi.org/10.1109/TLT.2011.23.

Aristotle. (2009). *The Nicomachean Ethics n/e*. Brown, L. (ed.) Ross, D. (Trans.) (Revised edition). OUP Oxford.

Bietti, E. (2020). From Ethics Washing to Ethics Bashing: A View on Tech Ethics from within Moral Philosophy. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20. *Association for Computing Machinery* (pp. 210–219).New York. https://doi.org/10.1145/3351095.3372860.

Boddington, P. (2017). *Towards a code of ethics for artificial intelligence research*. Berlin: Springer.

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-Centred learning analytics. *Journal of Learning Analytics, 6*(2), 1–9. https://doi.org/10.18608/jla.2019.62.1.

Bull, S., & Kay, J. (2016). SMILI☺: A framework for interfaces to learning data in open learner models, learning analytics and related fields. *International Journal of Artificial Intelligence in Education, 26*(1), 293–331. https://doi.org/10.1007/s40593-015-0090-8.

Conati, C., Porayska-Pomsta, K., & Mavrikis, M. (2018). AI in education needs interpretable machine learning: Lessons from open learner Modelling. *ArXiv:1807.00154 [Cs]*. http://arxiv.org/abs/1807.00154

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

DeepMind Ethics & Society. (2017). https://deepmind.com/about/ethics-and-society

Deshpande, K. V., Pan, S., & Foulds, J. R. (2020). Mitigating demographic Bias in AI-based resume filtering. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 268–275. https://doi.org/10.1145/3386392.3399569.

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue a checklist for trusted learning analytics. Proceedings of the sixth international conference on Learning Analytics & Knowledge, 89–98. http://dl.acm.org/citation.cfm?id=2883893

European Union. (2019). Ethics guidelines for trustworthy AI. Independent High-Level Expert Group on Artificial Intelligence. European Union. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., & Vuorikari, R. (2016). Research evidence on the use of learning analytics: Implications for education policy. http://oro.open.ac.uk/48173/

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

Friedman, B., & Kahn, P. H., Jr. (1992). Human agency and responsible computing: Implications for computer system design. *Journal of Systems and Software, 17*(1), 7–14.

Future of Life Institute. (2013). Future of Life Institute. https://futureoflife.org/ai-safety-research

Gearhart, D. (2012). Lack of Ethics for eLearning: Two sides of the ethical coin. *International Journal of Technoethics, 3*(4), 33–40. https://doi.org/10.4018/jte.2012100103.

Holmes, W., Bektik, D., Whitelock, D., & Woolf, B. P. (2018). Ethics in AIED: Who Cares? In C. P. Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay (Eds.), *Artificial Intelligence in Education* (Vol. 10948, pp. 551–553). Cham, Switzerland: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-93846-2.

Holmes, W., Bektik, D., Di Gennaro, M., Woolf, B. P., & Luckin, R. (2019a). Ethics in AIED: Who cares? In S. Isotani, E. Millán, A. Ogan, P. Hastings, & R. Luckin (Eds.), *Artificial Intelligence in Education* (Vol. 11625, pp. 424–425). Cham, Switzerland: Springer International Publishing AG. https://doi.org/10.1007/978-3-030-23204-7.

Holmes, W., Bialik, M., & Fadel, C. (2019b). Artificial intelligence in education. *Promises and Implications for Teaching and Learning.* Center for Curriculum Redesign.

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial Intelligence in Education* (Vol. 11625, pp. 157–171). Cham, Switzerland: Springer International Publishing AG. https://doi.org/10.1007/978-3-030-23204-7.

House of Lords. (2018). *Ai in the Uk: Ready, Willing and Able*. https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

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.

Kobsa, A. (2007). Privacy-enhanced web personalization. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web: Methods and Strategies of Web Personalization* (Vol. 4321, pp. 136–154). Berlin: Springer. https://doi.org/10.1007/978-3-540-72079-9_4 Generic User Modeling Systems.

Labarthe, H., Luengo, V., & Bouchet, F. (2018). Analyzing the relationships between learning analytics, educational data mining and AI for education. In Guin, N. & Kumar, A. (Eds.), *ITS 2018 Workshop Proceedings* (pp. 63–71).

Lewthwaite, S. (2014). Web accessibility standards and disability: Developing critical perspectives on accessibility. *Disability and Rehabilitation, 36*(16), 1375–1383. https://doi.org/10.3109/09638288.2014.938178

Macfarlane, B. (2003). *Teaching with integrity: The Ethics of higher education practice*. Abingdon: Routledge.

Mobasher, B., Kleanthous, S., Ekstrand, M., Berendt, B., Otterbacher, J., & Shulner Tal, A. (2020). *FairUMAP 2020: The 3rd Workshop on Fairness in User Modeling, Adaptation and Personalization. Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 404–405. https://doi.org/10.1145/3340631.3398671.

Murtarelli, G., Gregory, A., & Romenti, S. (2020). A conversation-based perspective for shaping ethical human–machine interactions: The particular challenge of chatbots. *Journal of Business Research*, S0148296320305944. https://doi.org/10.1016/j.jbusres.2020.09.018.

Norman, D. (2019). Why might human-centered design be harmful? Jnd.Org. https://jnd.org/human-centered_design_considered_harmful

Peters, R. S. (1970). *Ethics and education*. London: Allen & Unwin.

Porayska-Pomsta, K., & Rajendran, G. (2019). Accountability in human and artificial intelligence decision-making as the basis for diversity and educational inclusion. In J. Knox, Y. Wang, & M. Gallagher (Eds.), *Artificial Intelligence and Inclusive Education: Speculative Futures and Emerging Practices* (pp. 39–59). Singapore: Springer. https://doi.org/10.1007/978-981-13-8161-4_3.

Potgieter, I. (2020). Privacy concerns in educational data mining and learning analytics. *The International Review of Information Ethics*, *28*. https://informationethics.ca/index.php/irie/article/view/384

Reich, J., & Ito, M. (2017). From good intentions to real outcomes: Equity by design in learning technologies. *Digital Media and Learning Research Hub*.

Richards, D., & Dignum, V. (2019). Supporting and challenging learners through pedagogical agents: Addressing ethical issues through designing for value. *British Journal of Educational Technology, 50*(6), 2885–2901. https://doi.org/10.1111/bjet.12863.

Sacharidis, D., Mukamakuza, C. P., & Werthner, H. (2020). Fairness and diversity in social-based recommender systems. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 83–88. https://doi.org/10.1145/3386392.3397603.

Schelenz, L., Segal, A., & Gal, K. (2020). Best practices for transparency in machine generated personalization. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 23–28. https://doi.org/10.1145/3386392.3397593.

Schreck, J. (2003). *Security and Privacy in User Modeling* (Vol. 2). Netherlands: Springer. https://doi.org/10.1007/978-94-017-0377-2.

Sharkey, A. J. (2016). Should we welcome robot teachers? *Ethics and Information Technology, 4*(18), 283–297.

Slade, S., & Prinsloo, P. (2013). Learning analytics ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529.

Tarran, B. (2018). What can we learn from the Facebook—Cambridge Analytica scandal? *Significance, 15*(3), 4–5.

The Institute for Ethical AI & Machine Learning. (2018). https://ethical.institute

UNESCO. (2019). Artificial intelligence in education: Challenges and opportunities for sustainable development.

Université de Montréal. (2018). Declaration of Montréal for a responsible development of AI. Université de Montréal. https://www.montrealdeclaration-responsibleai.com

WEF. (2020). Chatbots RESET. A Framework for Governing Responsible Use of Conversational AI in Healthcare. World Economic Forum. http://www3.weforum.org/docs/WEF_Governance_of_Chatbots_in_Healthcare_2020.pdf

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., MyersWest, S., Richardson, R., Schultz, J., & Schwartz, O. (2018). *AI now report 2018*. AI Now Institute: New York University.

Winfield, A. F. T., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A, 376*(2133), 20180085. https://doi.org/10.1098/rsta.2018.0085.

World Economic Forum. (2019). *Generation AI.* May 2019 *Workshop Report.* http://www3.weforum.org/docs/WEF_Generation_AI_%20May_2019_Workshop_Report.pdf

Zapata-Rivera, J.-D., & Greer, J. E. (2004). Interacting with inspectable bayesian student models. *International Journal of Artificial Intelligence in Education, 14*(2), 127–163.

Affiliations

Wayne Holmes[1] · Kaska Porayska-Pomsta[1] · Ken Holstein[2] · Emma Sutherland[3] · Toby Baker[3] · Simon Buckingham Shum[4] · Olga C. Santos[5] · Mercedes T. Rodrigo[6] · Mutlu Cukurova[1] · Ig Ibert Bittencourt[7] · Kenneth R. Koedinger[2]

[1] UCL Knowledge Lab, UCL, London, UK

[2] Carnegie Mellon University, Pittsburgh, PA, USA

[3] Nesta, London, UK

[4] University of Technology Sydney, Sydney, New South Wales, Australia

[5] UNED, Madrid, Spain

[6] Ateneo de Manila University, Manila, Philippines

[7] Federal University of Alagoas, Maceió, Brazil