



Design of a future scenarios toolkit for an ethical implementation of artificial intelligence in education

Ana Mouta¹ · Eva María Torrecilla-Sánchez² · Ana María Pinto-Llorente²

Received: 6 April 2023 / Accepted: 21 September 2023 / Published online: 5 October 2023
© The Author(s) 2023

Abstract

In the 1970s, research on artificial intelligence in education emerged with the aim of acknowledging and accommodating the psychological aspects of the learning process. Since then, its applications have evolved and it is now used for student learning and assessment, teachers' pedagogical practice, management of educational institutions, and lifelong learning. Nevertheless, the ethical challenges of educational programmes using these systems have not been thoroughly studied. Anchored on the theoretical frame of dialogic ethics, this paper presents a section of a participatory futures research project. The goal of the research is to develop a toolkit that educators can use to ensure a smooth and ethical transition to artificial intelligence-based education while preserving the interests of educational development. This paper emphasises the need for an informed and participatory process that involves all stakeholders and begins with an expert consultation through the Delphi method, the results of which allowed the construction of eight hypothetical futures scenarios. These scenarios provide evidence that examining the ethics of using artificial intelligence systems presents an opportunity to reflect on the ethics of education as a whole. They highlight the challenge of balancing the benefits and drawbacks of such systems, especially concerning educational goals and the interplay between diverse educational actors and personal development in educational settings. The study outcomes are intended to encourage discussions on the integration of ethical artificial intelligence in education and facilitate the continuing professional development of teachers by equipping them with scenarios that can be used as a resource for training purposes.

Keywords Artificial intelligence in education · Ethics · Delphi method · Toolkit

✉ Ana Mouta
mouta.ana@usal.es

¹ Faculty of Education, University of Salamanca, P. de Canalejas, 169, Salamanca 37008, España

² Faculty of Education, University of Salamanca, P. de Canalejas, 169, Salamanca 37008, España

1 Introduction

Although artificial intelligence in education (AIED) research emerged in the 1970s, it has evolved slowly, and it is only in the last decade that the use of these systems has seen a boom in Western countries. A recent literature review of AIED technologies from 1993 to 2020 found evidence of usage for a variety of systems targeting school management, students, teachers and lifelong learning (Zhang & Aslan, 2021). These systems seem to hold promise for education as they are able to make decisions in complex situations, update behaviour in response to environmental changes, and coexist with other systems and people in physical environments (Dignum, 2021). According to Luckin et al. (2016), for teaching and learning artificial intelligence (AI) offer is wide-ranging and encompasses equipping teachers with AI teaching assistants, the provision of personalised support for each learner and individual tutors for learners in every subject. AI can analyse vast amounts of data about each student, including their learning styles, strengths, weaknesses, and preferences. This data-driven approach allows AI systems to create personalised learning paths and recommend specific learning materials, resources, and activities that align with each student's unique requirements and pace of learning (Miao et al., 2021). AIED may also help educators understand how learners are acquiring a wide range of skills. This is made possible through embedded assessments within the learning process, timely evaluations, the ability to adapt to learners' aptitude and knowledge levels, refreshed insights into learning progress, and the identification of factors influencing learning (U.S. Department of Education, Office of Educational Technology, 2023). This AI-driven adaptive assessments offer a more dynamic and responsive process to each student's performance. Adaptability ensures that students are neither overwhelmed nor bored by the assessment, as the difficulty level matches their current proficiency. By monitoring language patterns and interactions, AI systems can also gauge a student's level of engagement and emotional state during the learning process (Luckin et al., 2016). This information can help teachers adjust their pedagogical methods to better respond to this sentiment analysis. This way educators may help students to better understand and improve emotional preparedness for the educational process.

Intelligent tutoring support can identify areas where a student is struggling and offer targeted explanations, hints, and feedback to help the student grasp the material (Wang et al., 2023). The tutoring support can take various forms, such as interactive simulations, virtual dialogues, or step-by-step problem-solving walkthroughs. Moreover, intelligent tutoring systems can learn from the interactions with students and improve their effectiveness over time. The more data the AI system collects, the better it becomes at predicting the most effective approaches for individual learners, making the tutoring experience more efficient and relevant (Khosravi et al., 2022). By analysing historical data of student performance, AI systems can make informed predictions about future learning outcomes. This enables educators to identify students who might be at risk of falling behind or excelling, allowing for timely intervention and support.

By moving beyond the traditional 'stop-and-test' approach, AI-based education has the ability to address achievement gaps, enhance teacher proficiency, mitigate teacher turnover, and alleviate areas with significant teacher shortages (Luckin et

al., 2016). AIED is also capable of offering more intelligent and timely professional development tools, while also supporting parents in their efforts to assist their child's learning. Furthermore, AI systems can continuously update learning content based on emerging trends, new research, or changing educational standards. This ensures that students have access to the most current and relevant information, providing a more up-to-date and accurate learning experience. AI-based school management systems can automate administrative tasks such as resource allocation, scheduling, admissions, timetabling, attendance and homework monitoring, and school inspections (Miao et al., 2021). They can help optimise school operations, enhance efficiency, and improve communication between stakeholders. Finally, lifelong learning companions will be available to advise, recommend, and track learning, and more flexible learning environments will allow learners to study at their preferred time and location (Luckin et al., 2016).

But at what cost to personal and social development do these systems operate? Richards and Dignum (2019) affirm the need for a systematic examination of the values and ethics that justify the use of these technologies in education, considering the pedagogical approaches they can foster and their societal impact.

1.1 Challenges of using AI in learning environments

Targeting students and embedded in these systems, pedagogical agents were conceived two decades ago to attractively simulate human-like interactions between learners and content. Positive emotions, which benefit the learning experience and academic performance, are also incorporated as affective components in artificial agents (Dobrovestnova & Hannibal, 2020). Nonetheless, some concerns related to affective privacy, emotion induction, and virtual relationships between a human and an agent may arise from these interactions (Hudlicka, 2016). Social robots and talking dolls also have the potential to bring about changes on children's moral development (Williams et al., 2018). The Council of Europe (2022) has reflected on some other challenges arising from the use of AI-based systems in education: there is insufficient evidence for their effectiveness, their impact on cognition is still unknown, and these technologies seem to limit not only students' but also teachers' agency. There are concerns regarding systems made for teachers, which could result in automating ineffective pedagogical practices and disempowering teachers and parents alike. Similarly, implementing AI in school administration faces several challenges that need to be addressed. Using AI to learn about students through learning analytics raises privacy concerns (mood analysis and activity logs to hit political views, ethnic identity, health or sexual orientation), safety, trust and fairness (Tundrea, 2020). Utilitarianism and deontology are challenged since data collected is not only related with the learners, but with their colleagues and even family. Another aspect to consider is the different assumptions on AIED ethics among private organisations, developers, government agencies, research centres, universities and schools (Popenici & Kerr, 2017).

1.2 Coverage of AIED ethics by research, policy and training

Given the above challenges, the present study aims to develop a toolkit of scenarios to reflect on the ethics of education in the advent of AIED that can be used in teachers' continuing professional development. Despite the considerable attention given to the ethics of general AI through numerous studies, principles, and regulations (Jobin et al., 2019; Nguyen et al., 2022), there has been no research conducted on creating tools to enhance teachers' ability to effectively utilise this technology. Furthermore, systematic educational policies for AIED are still a mirage. In fact, countries like China, India, Italy, Kenya, Malta, Singapore, South Korea, Spain and the United States are debating AIED in their policies, but only five of them include it in the context of their AI policies (Schiff, 2021). In what concerns the preparation for the use of AI in schools, capacity building focuses primarily on its technical component and it is almost only aimed at secondary and tertiary education in computer science courses. In fact, there is still a lack of teacher and parent education, and limited training opportunities for the general public (Miao et al., 2021).

Previous ethical approaches to using AIED seem to have failed to address a crucial pedagogical concern in light of the challenges posed by AI. Once AIED relates to the application of AI technologies in learning environments, its ethics must plainly consider the ethics of education. This means it should encompass teacher expectations' ethics, resource and expertise allocation, gender and ethnic biases, behaviour and discipline, accuracy and validity of assessments, knowledge quality, teacher roles, power relations between teachers and students, and particular approaches to pedagogy, such as constructivism (Holmes et al., 2021). Additionally, in this process of AI integration into classrooms, it is expected that education will continue to serve as a space for the democratic formation of public thought, language, and concepts related to social, economic, political, cultural, ethical, and caring aspects of life (Lynch, 2022).

1.3 Arguments for a participatory approach to AIED use ethics

Recognising the significance of stakeholder involvement in the design of large societal projects (Bahadorestani et al., 2020), this study strives to give educational actors a voice in establishing a secure and valuable environment when using AIED technologies. The novelty of this research topic and the limited number of project impact assessments make it particularly relevant to involve educational actors in discussions. Zuboff (2019) highlights the exceptional nature of these technological advancements that cannot be captured by current frameworks, reinforcing the importance of using participatory methods and futures approaches. Stiegler (1998) characterises technology as the science that accompanies the creation of technical objects, and he believes that its greatest strength is precisely its unpredictability. Unlike what the “robot myth” suggests, technology's dynamics are not controlled by automation, but rather by the object itself that is prone to unpredictability, making it difficult to predict its future development. However, by using “technology maieutic” experts can contribute to a constructive evaluation of the future of these systems and their applications. Therefore, it is crucial to incorporate a variety of expertise, including local and tra-

ditional knowledge and practices, in policy design, implementation, and evaluation of AIED. This approach also increases transparency, accountability, and legitimacy of decision-making (ICAT, 2020). Furthermore, regarding technologies, few research articles have examined how stakeholder engagement is considered by research teams to evaluate key characteristics of the technologies to be developed (Nygaard et al., 2021). So, this is the proposal of the research presented throughout this paper: to merge ethics with participatory, deliberative and stakeholder approaches, based on the assumption that the public can make an ethically informed assessment of a new technology and that the moral insights of various individuals involved in the creation of these systems can enhance ethical evaluations (Brey, 2017). In fact, from a pragmatic point of view, a participatory study can lead to a more diverse and comprehensive analysis; and from an ethical perspective, this study recognises the human right to be part of public decision-making processes that affect people's lives. This idea aligns with the democratic and emancipatory ideals of the modernist Enlightenment (Santos, 2012). So, the implementation of the ethics of discussion is justified because it is deemed more suitable than the Kantian dialectic in identifying practical solutions to real-world ethical dilemmas. The ethics of discussion is based on a dialogical concept of reason inspired by the "linguistic turn" of analytical philosophy. The tasks of deontological ethics must be carried out by communicative reason, embodied in an open discussion with the plurality of members of an ideal community of argumentation. This would allow for a closer connection between ethical argumentation, thought, and practical action, insofar as the social agents themselves, as ethical subjects, participate in argumentative activity and introduce various ethical contents into the discussion in their materiality (Santos, 2012).

Therefore, this study intends to answer to the following research question: How do diverse stakeholders' ethical viewpoints regarding the integration of AIED impact the shaping of potential future scenarios? How can these imagined scenarios be effectively utilised to craft a continuing professional development toolkit that supports the ongoing growth of educators, specifically in addressing the ethical dimensions tied to the utilisation of AIED? Slaughter (2020) suggests that concepts regarding the future should be incorporated into the curriculum, teacher training, and educational systems. Thus, this study will further enable conversations with educators concerning the primary themes of a training curriculum, which will aid in the facilitation of the toolkit's use. The aim of this training programme is to guarantee that AIED is used in a meaningful and ethical manner that prioritises educational and pedagogical objectives. This toolkit can also drive both risk and impact assessments that support educational policy design and implementation.

As AIED theory is still in its early years of growth, this research integrates futures studies methodologies to anticipate, monitor, and address the ethical challenges that these technologies may pose (Gidley, 2017). For more than 60 years, futures studies have evolved from a method that makes predictions to a method that questions the possible, probable, or preferred transformations and impacts of an existing object as it moves to the future (Hines, 2020). The fourth approach for conducting these investigations involves participatory action learning/research, which centres around stakeholders taking an active role in shaping their own future, drawing on their beliefs about what the future holds and what is most important to them (Inayatul-

lah, 2007). Mapping alternative futures is also a way of agency looking into the future. So, the option of creating techno-ethical scenarios is justified by the fact that they have proven to be appropriate to study moral change. They allow for an ethical analysis based on the expected future moral values of the stakeholders involved (Brey, 2017). In particular for this research, it was chosen a participatory scenario planning approach (van Notten, 2006) and most of the content is based on the Delphi consultation (Bond et al., 2021; Dinges et al., 2020; Nuwan et al., 2021). Delphi is a useful futures studies method that enables idea generation on unexplored or controversial topics by bringing together anonymous specialists from diverse regions and disciplines, allowing freedom of expression and change of opinion (Green, 2014).

2 Methodology: the Delphi method

The main objective of this study is to investigate the ethical challenges related to the integration of AIED from the perspective of multiple educational stakeholders using the Delphi technique. The study aims to develop an informed toolkit that can be utilised in continuing professional development for educators in different regions. The expertise, perspectives, and viewpoints of experts were sought to gather insights on various aspects, such as AIED technologies, applications, purposes, contexts, educational actors, subjective experience, impact on subjectification, socialisation, and qualification, as well as usage drivers, ethical concerns, and existing regulations. The ultimate goal is to provide educators with the necessary sensitivity, knowledge and resources, empowering them to participate in constructive discussions and make informed and meaningful decisions concerning the ethical integration of AIED in various educational environments.

2.1 Expert group constitution

The research coordination group consisted of three researchers and the expert group was purposefully formed through a criterion sampling method (Patton, 1990). This implied the selection of participants based on predefined criteria that focused on their substantive knowledge of the problem under study (Ogbeifun et al., 2016). Furthermore, there is controversy over the use of the term “expert” and how to appropriately identify a professional as such (Hasson et al., 2000). Therefore, in the context of this research and given the novelty of the debate on the ethics of AIED use (in research, policy and training), there was an urgent need to define consensual criteria for what an “expert” can be. Based on other Delphi studies (Arteaga-Martínez et al., 2021; García et al., 2019), proven knowledge, extensive professional experience in the field of study, and sensitivity to scientific research (grounded on previous collaborations) were selected as preferred criteria. Furthermore, the participants’ professional diversity was appreciated since it brings varying perspectives from individuals in distinct fields (Renzi & Freitas, 2015). To manage the impact of non-acceptances, a larger number of individuals was initially invited, surpassing the preferred group size (Ogbeifun et al., 2016). So, during the initial stage, 30 AIED experts were chosen from different regions, and eventually, 18 of them consented to take part. Out of the

selected participants, five individuals did not respond to the invitation, and seven faced difficulties in fully engaging with the process due to their professional commitments and ultimately declined the invitation.

The eligibility criteria selected were the following: 1. work experience in the field of technology for education (EdTech) as (a) government advocate, opinion-maker, or supplier; (b) researcher; (c) specialist in implementation and evaluation of technologies in education; and (d) specialist in EdTech development; 2. professional experience (PE) in the field of over 10 years; 3. previous collaboration with academic research (PCR); 4. perception of the self as a specialist in education with technologies or EdTech. If any of the four criteria were not met, the potential participant was deemed ineligible. For the first criterion, participants were considered eligible if they had profiles (a), (b), (c), or (d), but it was not expected for all four profiles to be present in the same participant. Criteria three and four were implemented to ensure highly qualified panel members with a high level of expertise. With this purpose, the coefficient of expert confidence ($K = \frac{1}{2} (Kc + Ka)$) was added (Almenara & Osuna, 2013; Sanromà-Giménez et al., 2021). Kc is understood as the self-assessment knowledge coefficient on the topic (on a scale of 0 to 10) multiplied by 0.1. The argumentation coefficient (Ka) was determined based on the participant's involvement in previous research (criterion #3) and their years of professional experience: 1 for more than 30 years; 0.8 for 20 to 30 years; 0.5 for 10 to 20 years. The coefficient of expert competence has been 0.73. To reduce observer bias, the data recorded in each round was analysed by multiple observers from the lead research team: the three researchers attempted to ensure interrater reliability of the collected data.

Responses to the first round came from 18 participants (100% participation rate) whose sociodemographic and occupational profiles are presented in Tables 1 and 2. The mean age of the participants is 44.5 years (SD 7.42) and 16.6% identify as women and 83.3% as men. They are employed in various continents, encompassing countries like Portugal (PT), Timor, United Arab Emirates (UAE), and United States of America (USA). All have professional experience in the field of education and hold either a Master's or PhD degree. The experts' occupational field can be grouped as follows: corporate and business (61%), academic (22%), government (11%), non-profit and community-based (6%).

Table 1 Experts' sociodemographic profile per generic professional category

Delphi Experts' Profiles								
Generic Professional Profile	No. of Experts	Age Range			Gender		Working Country	
1. EdTech government advocate, opinion-maker, supplier	5	31–40	41–50	51–60	Men	Women	PT	UAE
		2	3	0	4	1	4	1
2. Researcher on EdTech	5	31–40	41–50	51–60	Men	Women	PT	Timor
		1	2	2	4	1	4	1
3. EdTech Specialist (implementation & evaluation)	3	31–40	41–50	51–60	Men	Women	PT	USA
		1	1	1	2	1	2	1
4. Specialist in EdTech development	5	31–40	41–50	51–60	Men	Women	PT	USA
		2	2	1	5	0	3	2

Table 2 Experts' professional profile

Expert No.	Generic Professional Category	Qualification	Professional category	Occupational Field	PCR	PE	K
1	EdTech Advocate	Master in Management	EdTech broker	Corporate and Business	Yes	20–30	0,7
2	EdTech Specialist (implementation & evaluation)	Doctor in Education	Government EdTech advisor/ School Dean	Government	Yes	30–40	1
3	Researcher on EdTech	Doctor in Psychology	Researcher/ Education Lecturer	Academic	Yes	20–30	0,85
4	Specialist in EdTech development	Doctor in Psychology	CEO in Data Business Analytics	Corporate and Business	Yes	40–50	0,85
5	EdTech Specialist (implementation & evaluation)	Master in Computer Science/ Engineering	Joint venture analyst in EdTech	Corporate and Business	Yes	20–30	0,8
6	EdTech Specialist (implementation & evaluation)	Doctor in Sociology	Education programmes consultant and CPD trainer	Non-profit and community-based	Yes	10–20	0,65
7	EdTech Advocate	Master in Education	Government EdTech advisor	Government	Yes	20–30	0,8
8	Researcher on EdTech	Doctor in Anthropology	Researcher/ Anthropology Lecturer	Academic	Yes	20–30	0,75
9	EdTech Advocate	Master in Education	EdTech broker	Corporate and Business	Yes	20–30	0,75
10	Specialist in EdTech development	Master in Electrical Engineering	Software developer	Corporate and Business	Yes	10–20	0,6
11	Researcher on EdTech	Doctor in Chemistry	Researcher/ Computer Science Lecturer	Academic	Yes	30–40	1
12	Specialist in EdTech development	Master in Computer Science/ Engineering	Software developer	Corporate and Business	Yes	10–20	0,55
13	EdTech Advocate	Master in Management	EdTech broker	Corporate and Business	Yes	20–30	0,7
14	Specialist in EdTech development	Master in Civil Engineering	Software developer	Corporate and Business	Yes	20–30	0,8
15	Specialist in EdTech development	Master in Electrical Engineering	Systems engineer	Corporate and Business	Yes	20–30	0,75
16	Researcher on EdTech	Doctor in Philosophy	Researcher/ Media and Communication Lecturer	Academic	Yes	30–40	1
17	EdTech Advocate	Master in Computer Science/ Engineering	Chair of the Board at EdTech company	Corporate and Business	Yes	10–20	0,6
18	Researcher on EdTech	Doctor in Education	Pedagogy R&D in business and CPD trainer	Corporate and Business	Yes	10–20	0,7

2.2 Rounds implementation

The implementation of the Delphi method involved three iteration loops, with a synthesis facilitated by the researcher's regular feedback and the comparison of the results with informed literature (Green, 2014). The various rounds included (1) answering a questionnaire; (2) reviewing first answers and select the most important critical points for each criterion; (3) vote on the new ideas to define a final list of criteria – this one would provide the inputs for constructing hypothetical scenarios that reflect the ethical challenges AIED poses; (4) discuss the plausibility of the scenarios, rewrite them and select those that better portray the ethical challenges of AIED.

2.2.1 Iteration 1

Participants were given an 8-item questionnaire (cf. <https://forms.gle/2CqBDsyy3p2n1jpE8>) to share their knowledge, vision, and opinion on the intersection of AIED and ethics. In the process of designing the questionnaire, the research team ensured its validity by drawing insights from various sources and grounding the questionnaire items on relevant literature, specifically codes for the responsible use of AI. The Artificial Intelligence's Ethics guidelines for trustworthy AI from the European Commission (2019) and Nesta's (2019) Map of the global AI governance landscape were instrumental in shaping the questionnaire, particularly questions 4, 5, and 6. These documents shared similar principles and recommendations, encompassing AI creation, function, and outcome stages. While incorporating these recommendations, it was noted that some guidelines were broad and lacked specific guidance for practical implementation in educational settings. To address this, questions 1, 2, 3, and 7 were derived from Holmes et al.'s seminal work (2019) on the promises and implications for teaching and learning of AIED. To ensure reliability, the questionnaire underwent a pilot test with a small group of individuals similar to the target participants, including an Edtech developer, an educational researcher on ICT, and an Edtech purchaser. Their thorough review of the questions helped identify any ambiguities or misunderstandings, leading to necessary adjustments to enhance clarity. Only one minor change related to language clarity was made in the last question. Furthermore, the team employed the test-retest reliability approach to assess the stability of responses over time. Participants were asked to freely answer the questions and then answer them again after three months to evaluate the consistency of responses on separate occasions before the questionnaire was sent to the experts. After collecting the experts' 18 responses, the coordination team condensed each meaning unit to identify broader categories, and descriptive statistics were used to determine the frequency of each category.

2.2.2 Iteration 2

In the second round, the 18 participants were presented with the results and asked to rank the importance of each category based on their personal views. Of the participants, 12 (67% participation rate) submitted their responses, which is still within the recommended range of 5 to 20 experts for qualitative research on a new topic

(Landeta, 1999). While the participants were aware of all study phases, the decrease in participation rate can be attributed to the demanding nature of this research phase, which occurred during the sudden pandemic-related restrictions and uncertainty in 2020. Content that was classified as medium-high and high was incorporated at this stage (representing challenging or very challenging issues), and categories with a sum of frequencies equal to or exceeding eight (more than half of the participants' votes) were retained for the third iteration.

2.2.3 Iteration 3

A. Content Relational Analysis.

Following the tradition of merging Delphi data with current literature, the third and final round combined the collected data with “The Ethical Framework for AI in Education” (The Institute for Ethical AI in Education, 2021), created to guide the design, procurement and application of AI on behalf of learners. The goals of the ethical framework and the opportunities and challenges of AIED found by the experts were consistent and therefore interrelated by two elements of the coordination team and reviewed by two others. The cognitive mapping presented in Table 3 was used for the experts to construct hypothetical scenarios based on these ethical categories and the possible outcomes of AIED implementation in different scenarios and from the perspective of diverse educational actors.

B. Hypothetical Scenarios Construction.

Scenarios can be either normative or exploratory. Normative scenarios show ways to achieve desirable outcomes, while exploratory scenarios explore potential developments, regardless of whether they are desirable (Kosow & Gaßner, 2008). In this study, we followed the basic steps of exploratory scenario planning proposed by Dean (2019). The first step was the (1) scoping phase, which involved defining the exercise's thematic coverage, stakeholders, and timeline. The (2) information-gathering phase analysed various data sources, including updated key reports like “The Ethical Framework for AI in Education”. The (3) trend and uncertainty analysis involved analysing possible future situations in terms of their likely impact and level of uncertainty. The principal investigators of this study performed this analysis, followed by the Delphi experts in the third iteration, as further described below. In the (4) scenario-building phase, the coordination team created eight hypothetical scenarios based on the experts' input. These scenarios were designed as short exploratory vignettes that presented a difficult-to-solve dilemma, following the orthogonal construction (Wright et al., 2014) and portraying one of four situations (the horizontal axis representing the degree of impact and the vertical axis representing the degree of uncertainty). They describe potential risks that may arise while striving to achieve eight out of the nine goals outlined in “The Ethical Framework for AI in Education”. The objective of managing administration and workload was excluded from this analysis as the experts' insights regarding opportunities and challenges did not

align with this category. Table 3 provides an example of the approach taken during this phase.

All of these scenarios were thoroughly examined by the experts. This exercise, which involves using key criteria to assess scenario quality, has a long tradition (Greeuw et al., 2000; Kreibich, 2007). Although scenarios are of a hypothetical nature, they are by no means arbitrary and must be evaluated according to criteria such as plausibility, consistency, comprehensibility and traceability, distinctness, transparency, degree of integration, and quality of reception (Kosow & Gaßner, 2008). In this study, participants were asked to give feedback on each scenario based on five criteria: (1) plausibility – whether it seems possible, (2) consistency – whether it makes sense logically, (3) comprehensibility – whether it is easy to understand, (4) relevance – whether it is relevant, and (5) distinctiveness – whether it is different from the others.

3 Results per round

3.1 Iteration 1

For each category created upon the participants' responses, frequencies were determined by descriptive statistics (cf. Table 4).

3.2 Iteration 2

Regarding the future of AIED, some of the initially proposed categories have been excluded. This happened with the items “Employment” and “Relational and Societal Factors”, which look at how education relates to impacts on other specific social layers. This appears to be true for both the negative perspectives – “Loss of interaction & detachment” – and the positive ones – “More humanistic causes, leisure and culture” (both from the category “Relational and Societal Factors”) or “Better matching people-education-jobs” (category: “Employment”). The most extreme views concentrate a smaller number of votes: for example, “Education dissolution” (3 votes for medium-high and none for high) or “Schools dissolution” (5 votes combining medium-high and high). The highest rated answers correspond to questions more directly related to academic achievement and the improvement of didactic resources: “Knowledge Management & Share”, “Learning Processes” and “Skills enhancement”. Looking at the critical positive/negative spectrum, four dimensions stand out positively, meaning they can have a noteworthy impact: “Instant data uploads on any topic” and “Personalised/Smart Learning Platforms and MOOC's” (both from the category “Knowledge Management & Share”), “Real time engagement and performance assessment and feedback” (category: “Learning Processes”), “Enhanced high-level processing” (category: “Skills enhancement”). All of them related to improving academic performance. On the negative or more challenging side, four dimensions stood out: “Education as a business”, “Information as commodity”, “Mainstream thinking and standardised behaviours”, and “Larger learning divide”. All of these points belong to the “Broader Implications” category, showing that these experts

Table 3 Delphi third round: tables with hypothetical scenarios sent to experts
 SCENARIO 1. Objective from “The Ethical Framework for AI in Education”: Achieving Educational Goals
AI should be used to achieve well-defined educational goals based on strong societal, educational or scientific evidence that this is for the benefit of learner.

	AIED Targets	AIED Function	AIED Outcomes	Schools support
Future with AIED	School Administrators and Government	Content Production	Education Quality	Product Procurement
Information as commodity	Responsibility for frameworks and assessment	Accountability for content selection and curation	Accelerated Individuation	Framework definition for procurement
Mainstream thinking and standardised behaviours			Reproduction of trainable skills	
Knowledge Management & Share			Human-human Interaction	
Instant data uploads on any topic			Challenge in terms of emotions and deep meaning	
Personalised/Smart Learning Platforms and MOOC’s				
Skills enhancement				
Enhanced high-level processing				

Table 3 (continued)

Risks	AIED Targets	AIED Function	AIED Outcomes	Schools support
<p>SCENARIO 1. Objective from “The Ethical Framework for AI in Education”: Achieving Educational Goals <i>AI should be used to achieve well-defined educational goals based on strong societal, educational or scientific evidence that this is for the benefit of learner.</i></p> <p>Future with AIED</p>				
<p>Hypothetical Scenario: May, 2033. Chloe is wakened by her robot friend Roy. Chloe checks her personalised learning chronogram for the week and starts her learning day in her garden. She activates her mindfulness chip and breathes in deeply. The program suggested a Steve Jobs’ hologram on entrepreneurship for the morning, followed by a virtual meeting with John and Mitchell to discuss best practices to run a sustainable out-of-the-earth business ten years in the future. When they meet, they quickly benchmark what has already been done in the field, using their extended brains (what once was Internet is now connected to neural processes through implanted chips). Her backpack was prepared by Roy for the afternoon, with lots of sensors and some acrylic paints. Chloe creates a handmade painting that is able to provide multiple experiences through sensors: for example, if a viewer were to touch the sun, they might feel heat. All her academic performances are automatically graded on 10 parameters, including engagement, creativity, critical thinking, and empathy. She will only have one brick-and-mortar learning experience this week with ten students – five of them will meet on that day for the first time. The capacity to build learning opportunities with new colleagues is a highly appreciated skill set. She was given the names of the new colleagues and after receiving her teachers’ feedback along with the automated grading, she is ready for her favourite part of the day. She automatically screens all the new colleagues’ social media profiles through her favourite filters: dating history (longer and shorter relationships), trips to other countries, preferred clothing styles, and personal physical enhancements already completed. Each one immediately receives a score that matches Chloe’s already recorded preferences. She celebrates with Roy the fact that she found a 17-year-old boy who had a particularly high matching score. If nothing else, at least academic compatibility and empathy will be certain. She can hardly wait for that face-to-face learning day!</p>				

seem to agree on more global negative impacts of using AIED, namely in terms of politics and asymmetries related to (the quality of) access to education, its instrumentalization for profit and the emergence of dominant standardised attitudes.

3.3 Iteration 3

To assess and redesign the eight scenarios, the experts' feedback was analysed and broken down into meaningful parts. For instance, one expert raised a consistency question about the 2nd scenario related to evaluation: “[isn't there] more fear of correcting errors immediately or wanting to respond to the review? Do you think that could happen? The immediacy can trigger action and in some cases trigger fear” (Expert No.3). These inputs were considered and the scenarios updated accordingly. Five scenarios were evaluated based on how easy they were to understand. On the positive side, six scenarios were considered highly plausible, and five were deemed relevant. However, in terms of plausibility, five scenarios were less convincing (scenarios 1, 3, 4, 6, and 8). Parts of five out of the eight scenarios were neither believable (scenarios 1, 2, 4, 5, and 8) nor consistent (scenarios 1, 2, 3, 5, and 7). The fifth scenario, which concerns privacy, raised concerns about potential opposition during teacher training. The experts recommended clarifying the distinctions between scenarios 3 and 5 and proposed a dialogue between scenarios 1 and 8. Scenarios 1 and 5 underwent significant changes. The last three scenarios (6, 7, and 8) received fewer comments, either because they were more consistent or because the experts were less likely to provide feedback after initial involvement. In fact, the literature suggests that the optimal number of scenarios should be between two and four or five for ease of manageability (Dean, 2019).

In summary, the experts' recommendations were focused on several areas, including: (1) increasing the trade-offs between good and bad outcomes in the scenarios; (2) adding more biographical details to the characters; (3) setting some of the dates further in the future; and (4) creating clearer distinctions between scenarios that examine fairness and the preservation of a privileged status by private schools. Additionally, two participants predicted that some scenarios may be difficult for teachers to discuss due to their futuristic nature or because they portray teachers as passive. Regarding comprehensibility, the experts' understanding of future outcomes and trends resulting from AIED growth is depicted in Fig. 1. Certain factors are emerging as trends (low uncertainty and high impact) in the economic (E), political (P), social (S), and technological (T) domains. Other aspects reflect critical uncertainties (high uncertainty and high impact), primarily in the social (S) domain. Examples of these critical uncertainties include the dominance of classification and labelling systems, the gamification of human experience, AI's influence over emotional expression, the power of corporations as a civilisational threat, the importance of human relationships, the possibility of parental rejection of AI without proper ethical oversight, and the need to preserve students' aesthetic development in the highly technical global context.

After analysing all the data, a final list of future scenarios was generated (cf. https://drive.google.com/file/d/1o6ToayuZ80Knj4R6QsWBbL7NW0D0aAm7/view?usp=share_link). This list will serve as a comprehensive resource for continuing professional development with teachers.

Table 4 Delphi first round results: frequencies per category

1. If you were asked to forecast the future (academic/social/emotional data mapping for employability, online learning, impact and ROI...) of Education with Artificial Intelligence what would you highlight?	FREQ
Broader Implications	
Education as business	1
Education dissolution	1
Information as commodity	1
Mainstream thinking and standardised behaviours	2
Larger learning divide	1
School dissolution	1
Trade-off: Privacy	1
Employment	
Better matching people-education-jobs	4
Knowledge Management & Share	
Instant data uploads on any topic	2
Personalised/Smart Learning Platforms and MOOC's	3
Learning Processes	
Better teacher-student relationships	1
Learning personalisation	1
Learning practices impact & predictability	2
Real time engagement and performance assessment and feedback	2
Teacher's routine tasks facilitation	2
Relational and Societal Factors	
Loss of Interaction & Detachment	2
More humanistic causes, leisure and culture	1
Skills enhancement	
Enhanced high-level processing	4
Strengthened role of emotional intelligence and social skills	2
2. From the applications of AI in formal Education (automated assessment or feedback, curated content, learning personalisation, virtual teaching assistance...), which of them should encompass a particular ethically sound design?	FREQ
Automated Assessment/Feedback	
Integration of cognitive, social, emotional and psychological inputs into automated assessment	7
Curated Content	
Consider bias, limits, and manipulation	2
Learning Personalisation	
Constraint of personal freedom and capacity of thought	2
Misplacement of personal singularities	2
Privacy (publishers, vendors, schools, students)	2
Virtual Teaching Assistance	
Respect for students' self-determination and autonomy in virtual teaching	2
3. Regarding the learner, the educator or the school institution, which AI target would be of more ethical concern to you?	FREQ
Learner	
Education systems not prepared to respond to kids will to know more about AI	1
Performance-oriented solutions - avoidance of kids' "bad" emotions (frustration, boredom...) and refusal of its developmental insight	2
Students' exposure to AI conception shortcomings	2
Students' vulnerability to abusive power	6

Table 4 (continued)

Schools	
Data processing, encryption and storage	2
Educational actors training	2
Protection of school as a pre-technological space for thinking and imagining	1
School Administrators and Government	
Responsibility for frameworks and assessment	1
Teachers	
Preparation and training	3
Risk of replacement	1
4. When thinking of AI applied to Education, if you had accountability for AI creation which issues (accountability, biases, diversity representation, explainability, fairness...) would you be more ethically apprehensive about?	FREQ
Accountability	
Aligning industry's goals with education systems and individuals	5
Reinforcing accountability first to enable AI development	1
Diversity representation	
Inclusiveness, equity and fairness	8
Representation of local strengths and needs	1
Technological and ontological diversity	1
Explainability	
Fostering critical users	1
Enhancing transparent processes despite interests	2
5. When thinking of AI applied to Education, if you had accountability for AI function which issues (cyber security, decision making processes, hacking, privacy control...) would you be more ethically apprehensive about?	FREQ
Cyber Security	
Hacking and exploitation	5
Privacy control	6
Unauthorised use of personal information	3
Content Production	
Accountability for content selection and curation	4
6. When thinking of AI applied to Education, if you had accountability for AI outcome which issues (automation, citizenship, human-human interaction, labour changes...) would you be more ethically apprehensive about?	FREQ
Automation	
Blindfolded processes	4
Education Quality	
Accelerated Individuation	1
Reproduction of trainable skills	1
Human-human Interaction	
Challenge in terms of emotions and deep meaning	3
Danger of virtual interaction replication	1
Distance between people	1
Interaction among community members	1
Political Citizenship	
Vulnerability for those who have lower self-determination	3
Training and labour changes	
Demanding labour market	2
Students' academic and career development	1

Table 4 (continued)

Vocational Training threatened	1
7. In terms of school support (product procurement, students’ privacy, teachers preparation...), which areas would you consider of priority care?	FREQ
Educational actors	
Communities’ awareness and preparation	1
Involvement of all educational actors in AI rules definition	1
Product Procurement	
Framework definition for procurement	2
Sorted by internal school stakeholders	2
Teachers Preparation	
Capable of guiding students in tech usefulness	1
Greater AI literacy, awareness, and questioning	10
8. Considering the topics under study, please add any note or commentary, if you wish to.	FREQ
AI and its applications cannot be understood as all the same	1
AI cannot replace Education as human relation	1
AI outcomes are political matters	1
Philosophical conundrum of ethics: utilitarian position (“the most good for the most people”)	1
VS. Kantian universal ethics	
Schools should not be wired	1
Tech innovation should be an act of imagination exerted by all	1

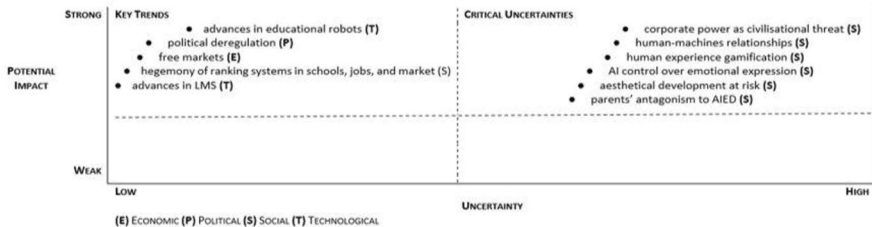


Fig. 1 Impact and Uncertainty grid for AIED futures scenarios

4 Discussion

As Holmes et al. (2021) suggest, the ethics of AIED is expected to call our attention to the ethics of education in the first place. Each new technology entering the realm of education is, in fact, an opportunity to rethink education ethics and how particular technological features may hinder specific aspects of pedagogy. Such challenging systems present an opportunity for schools to discuss and define their ethical common ground and to develop strategies to overcome any obstacles that may arise. Collaborating in this way can be helpful to identify and solve unexpected problems that may emerge while using AIED, problems that were not taken into account during the design process. So, which aspects of pedagogy need to be safeguarded by ethical considerations? By using the “ik.model” (Mouta et al., 2015), which is designed to assess how technology is integrated into education while prioritising educational goals, it is possible to grasp the potential risks and benefits of

having such an autonomous agent incorporated into pedagogy. First of all, it can be argued that the majority of ethical preoccupations of using AI in education were found in the “relational dimension” (the social domain of the impact and uncertainty grid), highlighting the ethical importance of how people interact with one another in the advent of AIED. The designed scenarios acknowledge the importance of shared values, including both explicit and implicit educational agreements. They also prioritise stakeholder involvement in decision-making processes, while being mindful of disparities in technological access and pedagogical quality. The first scenario centred around achieving educational goals and all participants agreed that the use of AI in education must be backed by robust evidence that demonstrate its beneficial effect on learners, what was previously highlighted by research (Richards & Dignum, 2019). The third scenario placed an emphasis on equity and highlighted the risk of a wider learning gap, where the disparities between public and private schools could become more prominent. This disparity extends to significant gaps between developed and developing nations, socioeconomic groups within countries, and those who have AI-enhanced jobs versus those who are susceptible to being replaced by them (Miao et al., 2021). The seventh scenario addresses informed engagement and recommends that students and other education actors should possess an adequate understanding of AI and its implications. The experts suggest that individuals with AIED knowledge and the ability to question should participate in establishing AI policies at the school level. Levinas (1969) contended that ethics must primarily acknowledge the importance of the interpersonal dimension, when in search of its existential ground. And this comes before any consideration of concepts like utility, virtue, or duty. Thus, it is crucial that the design, implementation (education with AIED), and evaluation (educational results) of activities with AIED be collaborative and shaped by people who have the capacity to consider the individual and societal benefits and drawbacks of its adoption and governance.

When designing these systems (“technological dimension”), it is important to be cautious about how the technologies are built, considering the type of connections people may form with machines. The sixth scenario emphasises the need for transparency and accountability in overseeing the operation of AI systems. In scenario eight, the importance of involving individuals who understand the potential consequences of AI on individuals and society in the design of these technologies is highlighted. This situation is exemplified by social robots that interact with humans. In fact, a review of the literature on the use of educational robots has evaluated their impact on four main dimensions, which are expected to be carefully scrutinised: (1) privacy; (2) human replacement; (3) impact on students; and (4) accountability (Serholt et al., 2017). To meet this need, it is encouraged to use a Trustworthy AI Ethics Guide in both creating and utilising AI technologies (European Commission, Directorate-General for Communications Networks, Content and Technology, 2019) and also to promote a participatory design of these technologies, informed by a variety of educational stakeholders and research fields.

Considering “content knowledge” and implementation through the lens of Levinas, education can be viewed as an ethical practice that aims to create spaces where individuals can engage with one another in a caring manner. The fourth scenario demonstrates how AI systems can undermine student autonomy, disregarding even

the most capable and perceptive students. Dependence on automated decisions and AI-driven personalisation can limit opportunities for student interaction and focus on knowledge that is easier to automate, hindering their development of resourcefulness, self-efficacy, self-regulation (Miao et al., 2021), and the recognition of themselves as the citizens they already are. Moral deskilling can also affect educators, who increasingly rely on AI machines to make decisions and become less critical and morally engaged (Tundrea, 2020). The fifth scenario focuses on privacy and the use of personal data to achieve educational goals. Experts warned of the possibility of education becoming a business with AIED being used, with many opportunities to enhance teacher training but at the expense of privacy and with the main goal of providing a specific service (Pammer-Schindler & Rosé, 2021).

In the dataism era, another ethical concern of AI-based education relates to the possibility of turning individuals into measurable and controllable entities through digital experiences. According to Han's (2014) argument that dataism could reduce self-tracking to mere self-surveillance, it's crucial to foster collaboration between teachers and students to envision and establish desirable futures with this unprecedented level of access to data. This is an invitation to reflect on what it means to be an individual in a group, and to foster mutual growth through reciprocal interactions. Educators also have the responsibility to unpack with their students the onto-epistemic grammar of dataism. This ethical undertaking involves exploring the anthropocentric perspective (Andreotti et al., 2015) underlying this desire, as well as the drive for ontological security (Lados et al., 2022) and the thirst for absolute knowability (Stein et al., 2017). This also provides a chance to use pedagogical strategies for a deeply purposeful and ethical learning experience. Project-based learning and curriculum infusion can be powerful strategies for achieving integrated goals. By incorporating other ethical, societal, and political concerns from different fields' perspectives, these approaches can address the challenges posed by systems that can grow in agency through our own inputs, while still meeting curriculum standards. Educators can use a variety of subjects and make students apply them to the task, while also availing themselves of AI resources. Engaging in discussions about AI functions from the perspectives of different subjects such as Mathematics, Science, History, and Languages (or any other) can serve as a means of strengthening newly acquired knowledge in these areas, applying it to practical and analytical tasks, and simultaneously building a more varied and intricate understanding of the AI systems in question. By doing so, education can move toward a more ethical exercise of freedom, even in the face of digital pressure.

Considering evaluation and "learning processes" dimension, AIED can provide just-in-time assessments, as well as new insights into how learning is progressing. But before recognising the potential benefits of incorporating AI-based assessment into learning environments, it is necessary to address ethical concerns related to educational assessment. While it is true that obtaining high-quality knowledge is extremely valued and that AI can improve the processes of encoding, storage, and retrieval by offering personalised pathways, discussions with education experts indicate that this can only be achieved if there is mutual agreement and respect between individuals and what to expect from their interaction with autonomous systems. This is even more crucial now, as the pandemic caused by Covid-19 has given new impe-

tus to technology (García-Peñalvo et al., 2021). As exam proctoring in some regions was a response to the problem of not being able to test students in physical situations, AI was identified as a possible solution to a large number of educational challenges. These ethical concerns were directed towards the second scenario that focused on forms of assessment. The experts believed that automated assessment and feedback on cognitive, social and emotional performance could become a reality in the near future and that this could present challenges and potential risks. Earlier research suggested that supervision is effective in decreasing deceitful actions. However, students may only behave honestly because they know they are being watched and not because of any intrinsic drive or self-reflection. This can lead to feelings of discomfort, such as a lack of privacy and anxiety, during the assessment process (Gudiño Paredes et al., 2021).

Drawing from Hannah Arendt's ideas, Coulter and Wiens (2002) suggest that in order to make sound educational evaluations it is essential to establish a connection between the teacher (actor) and the researcher (spectator). It is critical to challenge teachers to become accurate judges and actors themselves, which involves creating opportunities for them to appear. In fact, this is the goal and major ethical responsibility of this research: to engage teachers in the development of the curriculum for a course on AIED as part of their continuing professional development, using this scenarios toolkit as a basis for discussion. This represents an effort to urge teachers to become judging actors, which constitutes both a moral-political and an educational issue. These teachers are expected to engage with each individual child in complex communities, balancing guidance and agency and encouraging children to make informed judgments about the actions of others, and reflecting on their own actions and choices. In order to expect teachers to foster these skills, they must receive training in these very principles. Furthermore, the process of using AI systems to evaluate performance presents a challenge: everyone involved should be asked to participate in understanding the feedback given. When personal perceptions of performance differ from the classifications provided by the AI system, it can be both a valuable opportunity for personal growth (through insight) and understanding of how AIED works. This approach will allow for constructive criticism and questioning, forming the foundation for critical engagement with the world.

It's important to recognise that AI is more than just a neutral tool; it's an agent that learns, interacts, and can impact outcomes, which can create conflicts between students, teachers, and the educational system in terms of agency. While AI-powered chatbots and virtual assistants can provide students with 24/7 support and resources, thereby increasing their autonomy, there are also risks associated with AIED that could undermine this. For instance, if AI is used to make decisions such as determining which courses learners should take or what career path they should pursue, it could limit their options and opportunities for self-determination. This, in turn, may restrict their ability to explore their identity (which is crucial for psychosocial development) and form a sense of self. Similarly, if AI is used to monitor student behaviour or performance, it could lead to a surveillance culture that restricts students' ability to take risks and make mistakes, which are deemed essential for growth. Furthermore, the Vygotskian notion of 'scaffolding,' which involves a skilled mentor providing guidance and encouragement for action, may be interpreted differently in the context

of AI. Since AI may not be able to offer the same level of support and encouragement as a human mentor, it could alter the perception of the teacher's role, who is expected to provide challenging emotional experiences that are crucial for confidently engaging with the world.

To conclude, this toolkit aims to bring intervention by providing educators with a comprehensive set of resources and guidelines, enabling them to effectively address pedagogical challenges, including the integration of AIED as supportive tools when appropriate, considering ethical aspects and potential challenges. Designed for use in educator continuing professional development, the toolkit will consist of training modules and workshops covering fundamental AI concepts and their applications in education, how AI can address current educational needs, the benefits and challenges of AI implementation in the classroom, and the contextual integration of AI in various settings. It will also foster participatory and collective agency and decision-making among educational stakeholders to define ethical and pedagogical aspects of AIED implementation that better suit their educational contexts and interests. Additionally, this toolkit can be effectively employed as a scenario-based learning tool for students in project or inquiry-based learning, encouraging exploration of real-life situations and challenges that arise when using AI in the classroom, thereby empowering students and enhancing agency within the school environment. Finally, it can serve as a valuable resource for developers, providers, and educational decision-makers by offering guidance on ethical considerations related to AI usage in education. Within the scope of the present study, the subsequent stage will entail the concrete development and execution of a continuing professional development curriculum designed for educators. In this phase, a close partnership with teachers will be established to initiate the pilot testing of the toolkit within different educational environments. This methodological approach aims to elicit relevant insights, refine the toolkit's operational effectiveness, and systematically evaluate its influence on educators' pedagogical approaches and students' ethical learning experiences pertaining to AIED.

Some limitations of this study should now be highlighted. While this study's methodology may seem uncertain and speculative, it does not rely on predictive analysis, but rather on plausible or possible futures (Brey, 2017). One strength of the study is the use of short-term future narratives and providing information on the potential and dependencies of emerging technologies, which helps to bolster the decisions made throughout the work (Brey, 2017). Nevertheless, the study faces certain constraints, including the difficulty of conducting research grounded in objective moral reasoning, ensuring fairness, mitigating unequal power dynamics, and fostering equal participation (Hagendijk & Irwin, 2006). To mitigate these challenges, the Delphi method was used. Another limitation that should be acknowledged has to do with the fact this study involved a relatively small number of expert participants ($n=18$), which may raise concerns about the representativeness of the insights gathered. There is a risk that some perspectives or expertise relevant to the topic might be underrepresented, leading to conclusions that might not fully capture the complexity and nuances of the ethical challenges related to the integration of AIED in education. To address the limitations of the sample size, the research coordination team made a deliberate attempt to include a diverse group of experts from different regions, including Europe, Southeast Asia, Middle East, and North America. These experts possess var-

ied backgrounds and expertise in the field of AIED. There was also a concern to ensure that the eligibility criteria encompassed a range of professional profiles, such as government advocates, researchers, specialists in implementation, and specialists in EdTech development. Additionally, the Delphi method involves multiple iteration loops and expert feedback, meaning that data collection continues until a point of saturation is reached, where new insights or themes are no longer emerging from the panel. With 18 participants, it seemed possibly to efficiently reach this point, allowing for an in-depth exploration of the research questions. In the continuation of this research, qualitative research methods will be integrated, including focus groups with educators. These methods will complement the Delphi method, providing deeper insights into participants' perspectives and experiences.

5 Conclusions

This paper reports on a study that analysed an expert consultation on AIED. The goal of the study was to foster debate on ethical AI integration in education and support teachers' continuing professional development through scenarios that will serve as a toolkit for discussing training syllabi. The scenarios created feature a combination of current AIED technologies and some dystopian elements. They highlight how these systems may significantly impact our daily lives, interactions, thoughts, and emotions, being reasonable to expect that there may be significant challenges that arise at a societal or even civilisational level. Therefore, it's important for educators to be mindful of the potential risks and benefits of using AIED, particularly with regard to emotion recognition and social choice, and collaboratively establish purpose for its use. This means knowing fundamental characteristics, potentialities, and challenges of AI, including its general functions, and being transparent about how AI is being used. It also means involving students in decisions about how these technologies are being implemented, how its inputs are incorporated into the learning experience, acknowledging that AI should be used to support student agency.

The scenario toolkit created will serve as the foundation for conducting focus group discussions with educators, with the aim of anticipating the challenges and aligning educational objectives and practices with the context in which these AI technologies are intended to be employed. In this phase it will be determined how to integrate the data gathered into training programmes that promote the ethical use of AIED, while taking into account the diverse access, availability, and implementation scenarios across various regions. During these discussions, experts advised exercising caution, as teachers may feel uncomfortable discussing unfamiliar topics, causing the conversation to steer towards familiar territory. In the medium term, the goal of this research will be to equip educators with the appropriate resources to participate in such discussions, preventing resistance and fostering constructive dialogue that enhances the overall discourse.

The decision to employ a participatory method was taken to obtain a more comprehensive perspective on the ethical challenges of AIED implementation. Being the initial phase of a research project that will subsequently involve educators, this first step aimed to stimulate increased involvement from pertinent stakeholders who could

influence policymaking. The experts' unique considerations may already enable them to contribute to unprecedented critical evaluation of AI technologies' impact in education as responsible actors in the field. In addition, the methodology adopted in this study aims to conform to ethical principles, active participation and agency, which are the exactly same criteria proposed for the assessment of AIED. Not only does this methodology aim to ensure ethical research practices, but it also seeks to instil and promote the values it supports in the individuals involved, thereby guaranteeing their application in the way learners are encouraged to develop in the presence of AI-based systems.

Acknowledgements A portion of the data presented in this article was first introduced in a previous publication, “Blending machines, learning, sense of agency, and ethics: Designing an in-depth framework with experts using the Delphi Method approach” (<https://doi.org/10.1145/3486011.3486545>). This earlier paper provided a broad overview of the first author's PhD thesis, covering only the initial two iterations of the Delphi process and did not include the third iteration, which forms the core output of this current research. The present article offers a toolkit featuring eight AIED ethics scenarios, along with a more suitable introduction to the field of research, and a thoroughly description of the methodology used. In addition, it includes comprehensive sections for discussing and drawing conclusions from the research. In contrast, the previous version was primarily intended for academic purposes and formed a part of a PhD consortium, not providing a complete account of the research work.

Authors' contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ana Mouta, Eva María Torrecilla-Sánchez and Ana María Pinto-Llorente. The first draft of the manuscript was written by Ana Mouta and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The research leading to these results received funding from Fundação para a Ciência e Tecnologia under Grant Agreement No 2020.06258.BD. Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data Availability The datasets generated during and/or analysed during the current study are not publicly available to protect study participant privacy, but are available from the corresponding author on reasonable request.

Declarations

Competing interests All authors certify that they have no affiliations with or involvement in any organisation or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

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

- Almenara, J. C., & Osuna, J. B. (2013). La utilización del juicio de experto para la evaluación de TIC: el coeficiente de competencia experta. *Bordón*, 65(2), 25–38. <https://doi.org/10.13042/brp.2013.65202>.
- Andreotti, V., Stein, S., Ahenakew, C., & Hunt, D. (2015). Mapping interpretations of decolonization in the context of higher education. *Decolonization: Indigeneity Education & Society*, 4(1), 21–40.
- Arteaga-Martínez, B., Macías-Sánchez, J., Pla-Castells, M., & Ramírez-García, M. (2021). Design and validation of an observation instrument for mathematics classes in secondary education: Expert panel and Delphi process. *RELIEVE*, 27(2), art3. <https://doi.org/10.30827/relieve.v27i2.21812>.
- Bahadorestani, A., Naderpajouh, N., & Sadiq, R. (2020). Planning for sustainable stakeholder engagement based on the assessment of conflicting interests in projects. *Journal of Cleaner Production*, 242(46), <https://doi.org/10.1016/j.jclepro.2019.118402>.
- Bond, M., Marín, V. I., & Bedenlier, S. (2021). International collaboration in the field of Educational Research: A Delphi Study. *Journal of new Approaches in Education Research*, 10(2), 190–213. <https://doi.org/10.7821/naer.2021.7.614>.
- Brey, P. (2017). Ethics of Emerging Technologies. In S. O. Hansson (Ed.), *Methods for the Ethics of Technology*. Rowman and Littlefield International.
- Coulter, D., & Wiens, J. R. (2002). Educational Judgment: Linking the actor and the Spectator. *Educational Researcher*, 31(4), 15–25. <https://doi.org/10.3102/0013189X031004015>.
- Council of Europe (2022). Artificial intelligence and education: A critical view through the lens of human rights, democracy and the rule of law. <https://rm.coe.int/prems-092922-gbr-2517-ai-and-education-txt-16x24-web/1680a956e3>.
- Dean, M. (2019). Scenario Planning: A Literature Review. <https://doi.org/10.13140/RG.2.2.12629.24802>.
- U.S. Department of Education, Office of Educational Technology (2023). *Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations*. Washington, DC.
- Dignum, V. (2021). The role and challenges of education for responsible AI. *London Review of Education*, 19(1), 1, 1–11. <https://doi.org/10.14324/LRE.19.1.01>.
- Dinges, M., Wang, A., & Schuch, K. (2020). Using the Delphi method in evaluations: Incorporating a future oriented perspective in evaluations. *Fteval Journal for Research and Technology Policy Evaluation*, 50, 44–50. <https://doi.org/10.22163/fteval.2020.469>.
- Dobrosovstnova, A., & Hannibal, G. (2020). Teachers' Disappointment: Theoretical Perspective on the Inclusion of Ambivalent Emotions in Human-Robot Interactions in Education. In T. Belpaeme & J. Young (Eds.), *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 471–480). ACM. <https://doi.org/10.1145/3319502.3374816>.
- European, & Commission, Directorate-General for Communications Networks, Content and Technology, (2019). Ethics guidelines for trustworthy AI. <https://data.europa.eu/doi/10.2759/346720>.
- García, I. A., Moreno, M. F., Prado, G. R., Sáez, B., Puertas, M. A., & Gómez, A. A. (2019). Validación de un cuestionario sobre actitudes y práctica de actividad física y otros hábitos saludables mediante el método Delphi. *Revista Espanola De Salud Publica*, 93. <https://scielo.isciii.es/pdf/resp/v93/1135-5727-resp-93-e201909081.pdf>.
- García-Peñalvo, F. J., Corell, A., Abella-García, V., & Grande-De-Prado, M. (2021). Recommendations for Mandatory Online Assessment in Higher Education During the COVID-19 Pandemic. In D. Burgos, A. Tlili, & A. Tabacco (Eds.), *Radical Solutions for Education in a Crisis Context. COVID-19 as an Opportunity for Global Learning* (pp. 85–89). Springer Nature. https://doi.org/10.1007/978-981-15-7869-4_6.
- Gidley, J. M. (2017). *The future: A very short introduction*. Oxford University Press.
- Green, R. (2014). The Delphi technique in Educational Research. *SAGE Open*, 4(2), 1–8. <https://doi.org/10.1177/2158244014529773>.
- Greeuw, S. C. H., Asselt, M. B. A., Grosskurth, J., Storms, C., Rijkens-Klomp, N., Rothman, D. S., & Rotmans, J. (2000). *Cloudy Crystal Balls: An assessment of recent european and global scenario studies and models*. European Environment Agency. https://www.eea.europa.eu/publications/Environmental_issues_series_17.
- Gudiño Paredes, S., Jasso Peña, F. D., & de La Alcazar, F., J.M (2021). Remote proctored exams: Integrity assurance in online education? *Distance Education*, 42(2), 200–218. <https://doi.org/10.1080/01587919.2021.1910495>.
- Hagendijk, R., & Irwin, I. (2006). Public deliberation and governance: Engaging with Science and Technology in Contemporary Europe. *Minerva*, 44(2), 167–184. <https://doi.org/10.1007/s11024-006-0012-x>.

- Han, B. C. (2014). In M. S. Pereira (Ed.), *Psicopolítica*. Trans.). Relógio D'Água. (Original work published 2014).
- Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi Survey technique. *Journal of Advanced Nursing*, 32, 1008–1015. <https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x>.
- Hines, A. (2020). When did it start? Origin of the foresight field. *World Futures Review*, 12(1), 4–11. <https://doi.org/10.1177/1946756719889053>.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S., Santos, O., Rodrigo, M., Cukurova, M., Bittencourt, I., & Koedinger, K. (2021). Ethics of AI in education: Towards a community-wide Framework. *International Journal of Artificial Intelligence in Education*, 32. <https://doi.org/10.1007/s40593-021-00239-1>.
- Hudlicka, E. (2016). Virtual affective agents and therapeutic games. In D. D. Luxton (Ed.), *Artificial Intelligence in behavioral and Mental Health Care* (pp. 81–115). Elsevier.
- ICAT (Initiative for Climate Action Transparency) (2020). Stakeholder Participation Guide: Supporting Stakeholder Participation in Design, Implementation and Assessment of Policies and Actions. Climate, Community & Biodiversity Alliance and Verra. <https://climateactiontransparency.org/wp-content/uploads/2020/04/Stakeholder-Participation-Assessment-Guide.pdf>.
- Inayatullah, S. (2007). *Questioning the future: Methods and tools for Organizational and Societal Transformation*. Tamkang University.
- 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>.
- Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y. S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*, 3, 100074. <https://doi.org/10.1016/j.caeai.2022.100074>.
- Kosow, H., & Gaßner, R. (2008). *Methods of future and scenario analysis: Overview, assessment, and selection criteria*. German Development Institute. <https://www.idos-research.de/en/studies/article/methods-of-future-and-scenario-analysis-overview-assessment-and-selection-criteria/>.
- Kreibich, R. (2007). Wissenschaftsverständnis und Methodik der Zukunftsforschung. *Zeitschrift für Semiotik*, 29(2–3), 177–198.
- Lados, G., Hegedus, G., & Kovacs, Z. (2022). The role of identity and ontological (in)security in Return Migration: An empirical perspective from Hungary. *Journal of International Migration and Integration*, 24(2), 1–20. <https://doi.org/10.1007/s12134-022-00964-z>.
- Landeta, J. (1999). *El método Delphi. Una técnica de previsión para la incertidumbre*. Ariel.
- Levinas, E. (1969). *Totality and infinity: An essay on exteriority*. Duquesne University Press.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.
- Lynch, K. (2022). *Care and capitalism*. Polity Press.
- 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>.
- Mouta, A., Paulino, A., Ferreira, J., & Couto, F. (2015). Pedagogy as a Techné: Meaningful ICT Integration in formal Learning Scenarios. In G. Roman, J. Komuves & R. Hahn (Eds.), *Proceedings of the European Distance and E-Learning Network 2015 Annual Conference* (64–72). EDEN. <https://doi.org/10.38069/edenconf-2015-ac-0009>.
- Nesta (2019, January 22). Mapping global approaches to AI governance. <https://www.nesta.org.uk/blog/mapping-global-approaches-ai-governance/>.
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. P. T. (2022). Ethical principles for artificial intelligence in education. *Educational Technology Research and Development*, 1–25. <https://doi.org/10.1007/s10639-022-11316-w>.
- Nuwan, P. M. M. C., Perera, B. A. K. S., & Dewagoda, K. G. (2021). Development of Core Competencies of Construction Managers: The Effect of Training and Education. *Tech Know Learn*, 26, 945–984. <https://doi.org/10.1007/s10758-020-09474-2>.
- Nygaard, K., Graversgaard, M., Dalgaard, T., Jacobsen, B. H., & Schaper, S. (2021). The role of Stakeholder Engagement in developing New Technologies and Innovation for Nitrogen Reduction in Waters: A longitudinal study. *Water*, 13(22), 3313. <https://doi.org/10.3390/w13223313>. MDPI AG.

- Ogbeifun, E., Agwa-Ejon, J., Mbohwa, C., & Pretorius, J. H. (2016, March). The Delphi technique: A credible research methodology [Paper presentation]. 2016 International Conference on Industrial Engineering and Operations Management. Kuala Lumpur, Malaysia.
- Pammer-Schindler, V., & Rosé, C. (2021). Data-Related Ethics Issues in Technologies for Informal Professional Learning. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-021-00259-x>.
- Patton, M. (1990). *Qualitative evaluation and research methods*. Newbury Park.
- Popenici, S. A., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 22. <https://doi.org/10.1186/s41039-017-0062-8>.
- Renzi, A., & Freitas, S. (2015). The Delphi Method for future scenarios construction. *Procedia Manufacturing*, 3, 5785–5791. <https://doi.org/10.1016/j.promfg.2015.07.826>.
- Richards, D., & Dignum, V. (2019). Supporting and challenging learners through pedagogical agents: Addressing ethical issues through designing for values. *British Journal of Educational Technology*, 50(6), 2885–2901. <https://doi.org/10.1111/bjet.12863>.
- Sanromà-Giménez, M., Lázaro Cantabrana, J. L., Usart Rodríguez, M., & Gisbert-Cervera, M. (2021). Design and validation of an Assessment Tool for Educational Mobile Applications used with autistic learners. *Journal of New Approaches in Educational Research*, 10(1), 101–121. <https://doi.org/10.7821/naer.2021.1.574>.
- Santos, J. M. (2012). *Introdução à Ética Sistema Solar*.
- Schiff, D. (2021). Education for AI, not AI for education: The role of Education and Ethics in National AI policy strategies. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-021-00270-2>.
- Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S., & Paiva, A. (2017). The case of classroom robots: Teachers' deliberations on the ethical tensions. *AI & Society*, 32(4), 613–631. <https://doi.org/10.1007/s00146-016-0667-2>.
- Slaughter, R. A. (2020). Futures studies as a Quest for meaning. *World Futures Review*, 12(1), 26–39. <https://doi.org/10.1177/1946756719870277>.
- Stein, Z., Engberg, J., & Brown, T. (2017). The seductions of quantification: Measuring human rights, gender violence, and sex trafficking. *Signs: Journal of Women in Culture and Society*, 42(2), 397–422. <https://doi.org/10.1086/690937>.
- Stiegler, B. (1998). *Technics and Time, 1. The fault of Epimetheus*. Stanford University Press.
- The Institute for Ethical AI in Education (2021). The ethical Framework for AI in Education. <https://fb77c667c4d6e21c1e06.b-cdn.net/wp-content/uploads/2021/03/The-Institute-for-Ethical-AI-in-Education-The-Ethical-Framework-for-AI-in-Education.pdf>.
- Tundrea, E. (2020). Artificial Intelligence in Higher Education: challenges and opportunities. INTED2020 Proceedings, 2041–2049. <https://doi.org/10.21125/inted.2020.0644>.
- van Notten, P. (2006). Scenario Development: A typology of approaches. In OECD (Ed.), *Think scenarios, Rethink Education* (pp. 69–92). OECD Publishing.
- Wang, H., Tlili, A., Huang, R., Cai, Z., Li, M., Cheng, Z., Yang, D., Li, M., Zhu, X., & Fei, C. (2023). Examining the applications of intelligent tutoring systems in real educational contexts: A systematic literature review from the social experiment perspective. *Education and information technologies*, 1–36. Advance online publication. <https://doi.org/10.1007/s10639-022-11555-x>.
- Williams, R., Machado, C. V., Druga, S., Breazeal, C., & Maes, P. (2018). “My Doll Says It’s OK”: A Study of Children’s Conformity to a Talking Doll. In M. Hoiseith, G.E. Baykal, A. Peter, B. Farschian (Eds.), *Proceedings of the 17th ACM Conference on Interaction Design and Children* (pp. 625–631). Association for Computing Machinery. <https://doi.org/10.1145/3202185.3210788>.
- Wright, D., Finn, R., Gellert, R., Gutwirth, S., Schütz, P., Friedewald, M., Venier, S., & Mordini, E. (2014). Ethical dilemma scenarios and emerging technologies. *Technological Forecasting and Social Change*, 87, 325–336. <https://doi.org/10.1016/j.techfore.2013.12.008>.
- Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education Artificial Intelligence*, 2, 100025. <https://doi.org/10.1016/J.CAEAI.2021.100025>.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the New Frontier of Power*. Profile Books.