Acknowledgement

This version of the contribution has been accepted for publication, after peer review (when applicable) but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/[TO BE DETERMINED].

Citation

Please cite this manuscript as:

Borchers, C., Liu, X., Lee, H. H., Zhang, J. (2024). Ethical AIED and AIED Ethics: Toward Synergy Between AIED Research and Ethical Frameworks. Proceedings of the 25th International Conference on Artificial Intelligence in Education (AIED) — BlueSky Track. Recife, Brazil.

Ethical AIED and AIED Ethics: Toward Synergy Between AIED Research and Ethical Frameworks

Conrad Borchers[1(✉)][0000-0003-3437-8979], Xinman Liu[2][0009-0009-2903-0346], Hakeoung Hannah Lee[3][0000-0002-0567-7710], and Jiayi Zhang[4][0000-0002-7334-4256]

[1]Carnegie Mellon University, Pittsburgh, PA 15213, USA
cborcher@cs.cmu.edu

[2]University of Cambridge, Cambridge, UK
xl505@cam.ac.uk

[3]The University of Texas at Austin, Austin, TX 78712, USA
hklee@utexas.edu

[4]University of Pennsylvania, Philadelphia, PA 19104, USA
joycez@upenn.edu

Abstract. Ethical issues matter for artificial intelligence in education (AIED). Simultaneously, there is a gap between fundamental ethical critiques of AIED research goals and research practices doing ethical good. This article discusses the divide between AIED ethics (i.e., critical social science lenses) and ethical AIED (i.e., methodologies to achieve ethical goals). This discussion contributes paths toward informing AIED research through its fundamental critiques, including improving researcher reflexivity in developing AIED tools, describing desirable futures for AIED through co-design with marginalized voices, and evaluation methods that merge quantitative measurement of ethical soundness with co-design methods. Prioritizing a synthesis between AIED ethics and ethical AIED could make our research community more resilient in the face of rapidly advancing technology and artificial intelligence, threatening public interest and trust in AIED systems. Overall, the discussion concludes that prioritizing collaboration with marginalized stakeholders for designing AIED systems while critically examining our definitions of representation and fairness will likely strengthen our research community.

Keywords: ethics, bias, fairness, equity, representation, design, justice, reflexivity

1 Situating AIED Ethics and Ethical AIED

The field of artificial intelligence in education (AIED) increasingly embraces emergent ethical issues of artificial intelligence (AI). In a landmark paper by Holmes et al. [1], the field recognized the need to consider unintended implications of its technology regarding fairness, bias, equity, and representation. The authors delineate a gap between *doing ethical things* and *doing things ethically*, wherein there is a potential mismatch between the ethical implications of research goals (ethical things) and ethical practices within potentially unethical research goals (ethical doing). Specifically, current AIED research practices are predominantly concerned about "doing things ethically"—the technical and procedural validity (e.g., minimizing bias,

improving efficiency, prioritizing privacy) of deploying AIED systems, sidestepping the matter of "doing ethical things"—the normative validity concerning the purpose of an AIED system and whether it is in itself an ethical pursuit [1]. This mismatch begs the question: how can AIED productively integrate more fundamental critiques of its research with its practices? For the purposes of this study, we contrast critical and theoretical work on ethics in AIED (*AIED ethics*) and with AIED research practices aimed at promoting ethical good or avoiding ethical bad (*ethical AIED*). Prior work on general AI ethics has argued that ethical principles and frameworks are "toothless and useless" as they are isolated from practices and are not consequential [2]. However, counterexamples exist, where critical policy research collaborates with technical AI research to develop guidelines for auditing AI-based models, such as emerging large-language models, with implications for regulation and technical model audits at the core of research practice [3]. How may AIED achieve a similar synthesis to strengthen its ethical research practices? The present study contributes paths forward to synergistically integrate these differences in future AIED research.

AIED ethics and ethical AIED follow different definitions of ethics, with AIED ethics being *justice-oriented* and ethical AIED being *measurement-oriented*. AIED ethics stems from ethnographic and critical traditions in the humanities. It offers various *frameworks* that research and policy can learn from to enhance ethical AIED, with many calling for a more fundamental shift in research goals in AIED toward justice (e.g., prioritizing equity-enhancing design over technological advancements [4]). In contrast, ethical AIED research employs quantitative methods, increasingly aiming to *measure* issues such as bias and fairness in AIED systems [5]. While gaps between both approaches have been noted in critical algorithm studies [6], the present study describes how both discourses contrast in terms of specific ethical topics of interest in current AIED research: personalization, equity, representation, and bias. For the former, we describe contemporary methodologies to approach, measure, and improve each issue. For the latter, we summarize common definitions and critiques of each issue regarding AIED research and its ethical implications. Through this contrast, we synthesize *paths forward* of how AIED as a research field can produce practice and output that is not only ethical doing (ethical AIED) but also accomplishes ethical research goals (AIED ethics). We do not aim to draw a pessimistic stance that argues that both lenses are mutually exclusive. On the contrary, we describe recent examples of research that move the field toward achieving such synthesis and discuss what might be on the horizon for AIED researchers when such synthesis is achieved. Prioritizing such synthesis could make the AIED research community more resilient in the face of rapidly advancing technology and AI, which may threaten public interest and trust in AIED systems.

2 Evidence for a Disconnect between AIED Ethics and Ethical AIED

In discussing AIED ethics, we acknowledge that there are multiple traditions of ethical thinking related to AIED. For instance, Fox [7] delineates four traditions of consequential, ecological, relational, and deontological ethics in AIED research. Holmes and colleagues [8] have also advocated for a rights-based approach to AIED ethics. In recognizing ethical pluralism, the present paper broadly takes on a

justice-oriented lens as encompassing the recent discourse in critical studies of AIED ethics. This choice does not imply other ethical lenses on AIED are irrelevant; rather, it favors context-specific approaches to ethical justification that acknowledge the legitimacy of multi-stakeholder grounded realities [9]. As we will argue, a justice-oriented lens to AIED ethics, as opposed to incoherent sets of "toothless and useless" ethical principles detached from real practices, promises more synergies with ethical AIED and paths forward to improving AIED research practices [2].

Past positions of justice-oriented AIED ethics on AIED research practice can be broadly characterized as ones criticizing *measurable ethics*. Biesta [10] contests that contemporary education systems neglect the normative validity of educational measurements (i.e., what *should* be valued and thereby measured). In other words, focusing on the efficiency and effectiveness of learning processes sidesteps the normative questions of what defines a good education in the first place—the aims and ends for which these processes are directed. Commonly voiced concerns over AIED research goals today still echo this critique of measurement and effectiveness: intelligent tutoring systems prioritizing learning efficiency at the expense of collaborative and social interactions [11] or the datafication of student and teacher subjects potentially disintegrating into surveillance [12].

Taking ethical concerns seriously has been increasingly central in AIED, alongside the recognition that addressing ethical concerns upfront makes our community more resilient in the age of rapid technological progress. Still, awareness and debates about the ethics of AIED as a research field were not spotlighted until relatively recently [1]. This nascent discourse centers on computational approaches to boosting fairness and doing ethical good without much accounting for the *ethics of education*—that is, the purpose of learning, choice of pedagogy, human-computer relations, and access to education [1]. Contrasting these computational approaches, AIED ethics recognizes that data and algorithmic systems do not pre-exist the social actors, techno-scientific practices, institutional applications, and power struggles that bring them into being [13]. As such, AIED innovation should be understood as "a knot of social, political, economic and cultural agendas that is riddled with complications, contradictions and conflicts" [14, p. 6]. Can both lenses be integrated? In the following, we discuss how both AIED ethics and ethical AIED approach emerging areas of interest in AIED. After summarizing emerging AIED methodologies and discourse around each issue, we describe their discourse through AIED ethics, as signified by headers beginning with "beyond."

2.1 Four Issues of Ethical Discourse In and Around AIED

2.1.1 Issue 1: Personalization

Through the use of AI, personalization adopted in AIED systems presents an opportunity to provide high-quality and equitable learning access to students at scale: an ethical good. It is often modeled after human tutoring, analyzing students' needs and delivering tailored instructions and adaptive feedback [15]. This approach allows broader access to high-quality instruction, potentially enabling equitable learning opportunities for larger populations. It enables students to learn and progress without

being held back or left behind, which can often happen in a traditional classroom where instruction is standardized and delivered based on a fixed schedule [16]. The use of personalized instruction and learning pathways has been found to be beneficial in improving student engagement and learning [15].

Beyond Personalization: AIED ethics can guide thinking beyond the boundaries of typical personalization in AIED learning environments. While personalization seeks to deliver effective educational experiences to learners, it is often limited to micro-level decisions that offer learners individualized contexts, pacing, groupings, and pathways through prearranged materials. These personalized AIED technologies have been criticized for operating on behaviorist and instructionist pedagogies underpinned by a narrow understanding of personalization where "the pathway may be personalized but not the destination" [8, p. 34]. As such, learner agency is pre-determined and constrained within a set of universally standardized outcomes. Real personalization (according to AIED ethics)—or "subjectification" and "individuation" in Biesta's terms—involves cultivating learners' autonomy and capabilities to self-actualize and achieve what they individually want to achieve [10].

2.1.2 **Issue 2: Equality and Equity**

A central AIED research goal is to create systems that work equally well for different groups of learners. A common concern related to the issue of equity is the so-called "EdTech Matthew Effect," where AIED learning systems specifically benefit learners with high prior knowledge, deepening existing achievement gaps [17]. Equity relates to a constant relationship between effort and learning in AIED learning environments, such as learning opportunities in intelligent tutoring systems and learning gain. Recent research argued that learning rates (i.e., the average improvement in student accuracy per completed problem-solving step with feedback) are highly regular across students and within various learning domains [18]. Equality, or the absence of achievement gaps, could then be achieved if disadvantaged learners receive more learning opportunities in well-designed AIED systems. To evaluate and promote equity, AIED research has called for increasing use of school-level demographic variables to compare how different behaviors in AIED learning environments (e.g., help-seeking) differentially relate to learning outcomes across populations [19].

Beyond Equality and Equity: AIED ethics points attention to the fact that most AIED learning technologies are designed to intervene: they measure when a student struggles or does not reach a certain level of attainment. One risk of this deficit view is that realities of educational experiences could be masked by quantifiable participation and completion rates set by dominant institutions and regimes. AIED ethics cautions about leaving the measurement of what learners are lacking unquestioned. Why might learners be better off if AIED systems are mindful of setting educational goals? Interventions focusing on measurable outcomes overlook the underlying reasons necessitating additional support. For example, struggling to learn might relate to a range of cognitive and motivational factors. While predictive models can discern a subset of different sources of struggle [20], human teachers are often in a better position to diagnose learner needs, in line with a model of human-AI complementary permeating recent AIED successes [21]. However, learner needs

might have cultural roots that are left for critical AIED research to examine and design around. For instance, even when help is sought during the usage of intelligent tutoring systems, help-seeking behaviors vary across different sociocultural contexts, influenced by socioeconomic status, religion, power dynamics, and degree of individualism or collectivism from classroom to national levels of culture [22]. This variability underscores that the modeling and assessing these behaviors cannot be universally applicable. Similarly, learner access to AIED systems is rarely considered in tool design, with few notable exceptions [23]. These considerations underscore that more holistic considerations of learner contexts and obstacles to learning could improve learning environments and their ethical potential to reduce inequality [12].

2.1.3 **Issue 3: Representation**

Representation in ethical AIED focuses on integrating perspectives of users in the design, development, deployment, and evaluation of educational technologies and ensuring diversity encoded in demographic markers in datasets. This includes co-design practices involving educators, students, and other stakeholders in the creation of AIED systems [21]. Collaboration between researchers and stakeholders can potentially increase impact through more effective implementation and use. Research has also called for including diverse voices in the development process to ensure that AIED systems are reflective of a wide range of learning contexts and the needs of different learner populations [19]. By prioritizing representation, developers can mitigate the risk of perpetuating existing biases and ensure that technologies are inclusive and beneficial to a broad spectrum of users. Regarding the deployment and evaluation of AIED systems, past work has identified the need to study AIED systems in diverse cultural contexts and study their efficacy [23]. Recent research calls for the systematic study of AIED tools through the lens of the demographic groups that use them, which can be (among others) inferred from census data of schools AIED studies are run in, as student-level demographic data collection is often not feasible [19].

Beyond Representation: While ethical AIED highlights evaluating the effectiveness of AIED systems across social and cultural contexts [8, 24], AIED ethics emphasizes a critical attitude towards constructing and measuring learner categories. The most immediate approach to improve representation is to add, combine, and overlap identity categories such as class, race, gender, sexuality, ethnicity, ability, nationality, and age. However, AIED ethics argues that this approach leaves institutional forms of racialized, classed, gendered processes perpetuated by the dominant regimes of power unexamined. An additive approach to representation may be reductionist in serving as "a palliative to keep marginalized groups… from rebelling against a system that promotes structural inequality" [24, p. 25]. Accordingly, critically examining how AIED classifies learners (e.g., in terms of race, gender, or class) could improve the benefits of AIED systems for marginalized and underrepresented learners [25, 27]. Past research offers examples of this issue in education: without rethinking gender binaries within classification systems, we risk marginalizing non-binary learners [26]. Similarly, it is worthwhile to reflect on how alternative demarcations of race and ethnicity could serve underrepresented learners beyond North American contexts better. Within AIED, the construction of classification systems can constrain learners by overlooking within-group differences. For instance, one study argued that categorizing Laotian, Cambodian, and Vietnamese

American students within the broad "model minority" stereotype associated with East Asian Americans from China, Japan, and Korea ignores crucial within-group differences—such as academic performance and family resources—that are more predictive of educational outcomes than broader racial group distinctions [25].

How can critical lenses on representation inform AIED co-design practices? Design justice and liberatory philosophies that center on community-led and co-constructive practices [27] increasingly allow researchers to adopt participatory approaches and listen to marginalized voices among educational stakeholders [28]. Participatory approaches including diverse stakeholders in the design process) are compatible with current AIED co-design practices. However, through a critical lens, AIED ethics emphasizes a deliberate analysis of power within individuals, institutes, and where they intersect [25]. Consequently, they focus co-design efforts on listening to marginalized voices by establishing design spaces where marginalized groups can easily participate and envision alternative designs to current solutions [27]. Beyond attempts to measure representational fairness through categories [19], this approach can present opportunities for AIED tool design otherwise invisible to researchers.

2.1.4 **Issue 4: Bias and Fairness**

Student modeling in AIED systems involves using log data to model student behaviors and predict learning outcomes. Bias and fairness in these models are especially investigated from an algorithmic standpoint in emerging ethical AIED research. Algorithmic bias or fairness refers to the collective effort of examining the performance of student models, ensuring that the models are capable of providing unbiased evaluation for all learners regardless of their attributes [5, 29]. Algorithmic bias describes the problem where a data-driven predictive model functions better for some populations than others, producing disparate and poorer impacts for historically underrepresented or protected groups [29]. As predictions are often used to inform decisions and actions, algorithmic bias in a model can cause unfairness in the allocation of resources and misplacement of treatment. An increasing number of works in the field of AIED have dedicated their efforts to evaluating and improving the fairness of student models. Among them, a range of models have been examined across different student populations and intersectional groups [30]. To improve model fairness, studies suggested increasing data collection for minority students [31] and being critical with the decisions on the inclusion of demographic data [19].

Beyond Bias and Fairness: Efforts to improve algorithmic fairness in education assume the general benefit of innovation and aim to distribute these benefits equitably; in other words, no groups of learners should systematically benefit more from technology than others. AIED ethics asks about broader considerations of whether these approaches promote more ethical good or put historically disadvantaged learner groups at risk through second-order effects of technology. For example, while tools to profile learners may help prevent poverty-stricken students from dropping out of school, their intrusiveness can also undermine learners' rights to privacy and be misused by governing entities to distribute and rescind welfare. This is the case for Brazil's Bolsa Familia program (a direct income transfer program aimed at helping families out of poverty), where the use of facial recognition technology in public schools may lead to punitive consequences for families dependent on welfare

linked to monitored school attendance [32]. Such a program describes the ethical dilemma in algorithmic attempts to measure and enhance fairness in education requiring demographic data collection. AIED ethics highlights that these processes may heighten the visibility and, thereby, the vulnerability of historically low-income and marginalized groups. Eubanks [4] argues that while the expansion of digital systems in criminal justice, welfare, and education has increased the visibility of working-class women seeking public assistance, these systems also exacerbated their physical and economic vulnerability through behavioral surveillance and discoveries that would have gone unnoticed in the privacy afforded by wealthier families. Similarly, researchers warned against the danger of algorithms in increasing the vulnerability of already marginalized learners through further stigmatization [37].

What does AIED ethics suggest AIED research could do better about historical biases in present-day data? AIED ethics lenses on bias emphasize limitations of the promises of data neutrality and objective calculations in model-based approaches to bias. AI is not an inherently neutral set of technologies but rather embedded in and produced from human-run systems where historical biases are often entangled and untraceable [33]. As such, beyond improving accuracy and eliminating bias, AIED researchers and attempts should foreground their own positionality and reflexivity, including premeditation of how systems and algorithms they develop could be harmful to vulnerable learner populations. Researchers should "start from interrogating the existing inequalities, reflect their own position in the system of these inequalities and actively ask which constituencies will or will not benefit" [34, p. 331] rather than construing their personal involvement (e.g., motivations, beliefs, roles) in data protection as bad or biased practices. In AIED, special attention can be paid to the power disparities between those initiating and those subjected to AIED interventions [12]. We acknowledge that not all potential harm can be preemptively detected and no research to promote equality and support vulnerable populations is "risk-free." Still, AIED systems (including their public perception and likelihood of doing ethical good) could benefit from a holistic assessment of learner contexts and potential intervention risks in these populations.

3 Examples of Research Bridging AIED Ethics and Ethical AIED

Having delineated differences in how AIED ethics (i.e., justice-oriented, framework-heavy critical social science work) and ethical AIED (i.e., measurement-oriented, quantitative research practices) take on emerging challenges in AIED, the present study is not intended to paint a picture of insurmountable divides between both approaches. Rather, to aim at a synthesis of both approaches, we briefly summarize two example directions of successful synergies between both strands: a) research bridging AIED stakeholders and researchers and b) research reflecting on AIED research goals, practices, and conditions.

First, research bridging AIED stakeholders (e.g., students, teachers, caregivers) and researchers could promote ethical good and advancements in AIED. Practices that involve listening and designing around learner and educator needs are not only expected to amplify underrepresented voices but also lead to more effective AIED systems by listening to "weak signals" [35]. Studying AIED systems in the context of closely working with AIED stakeholders and studying the adoption of such

systems through observational methods also bears the potential of mitigating some of the concerns summarized around AIED ethics discourse earlier. First, studying where AIED systems break, fail, or do not help learners and why can address more fundamental issues around bias. Second, studying cultural practices beyond monitoring demographic data can reveal limitations in contemporary systems of classifying learners (representation). Third, studying adoption and system use beyond short-term studies could discover the potential harms of advanced technology on vulnerable populations (as in the Bolsa Familia program [32]). Fourth, designing around learner needs and concerns could potentially support learner self-actualization by setting broader learning goals beyond personalization within set learning goals. For research practice, observational methods could be supported by co-designing AIED tools with teachers [21] or specifically listening to underrepresented groups for whom existing AIED tools might not work well [27]. These approaches also focus on improving adoption before technology rollout, aligning the visions and needs of stakeholders with AIED tool development [36]. To make this vision of stakeholder involvement an actuality, AIED as a research community could support the creation of spaces where AIED stakeholders meet and work with researchers. A benefit of spaces for meeting and communicating can be an increase in trust in and adoption of AIED systems [12, 37]. Communication channels are expected to deepen trust between AIED stakeholders and research. To involve minorities in research more, efforts could focus on issues that research found relevant to trust in AI systems, for example, building consensus and best practices around data privacy standards [37].

The second example of bridging ethical AIED and AIED ethics is research reflecting on the positive and negative impacts of AIED research goals, practices, and conditions coming from within the AIED community. For example, recent work has qualitatively studied how the presence of analytics in AIED systems can introduce tensions in student mentoring relationships in higher education [12]. Specifically, the study noted that discrepancies between reported activity and data in activity reports could undermine trusted relationships between mentors and mentees. Therefore, the study calls for increased research on how AIED systems transform existing practices (e.g., in the classroom) and how participants perceive these systems rather than studying outcomes and fairness metrics based on performance alone. Similarly, overview articles reflecting on research paradigms and assumptions in AIED research can steer conversations around methodologies and approaches to measuring how AIED systems and research transform learners' lived experiences and promote ethical good. This includes position articles: a recent article argued that the majority of AIED systems operate on a deficiency-based lens, where intervention is provided to students who are lacking, at-risk, or not learning well, which underappreciates opportunities to design adaptivity around learner strengths and assets [38].

4 Paths Forward

Synthesized from the issues above and their differences when viewed from AIED ethics frameworks and ethical AIED research practice, we suggest paths forward that could promote the synthesis of both lenses.

4.1 **Path 1: Researcher Reflexivity**

Ethical AIED frameworks suggest that researchers question their research goals, definitions of issues such as bias, and the demographic lines along which learners are represented and studied. How can this lens be productively integrated into AIED research practice? Practically speaking, how does one go from theory to conceptualization to practice? One approach could be to carefully evaluate and communicate the feasible expectations and limitations of AIED systems to those involved and affected (e.g., learners, teachers, caregivers). Practicing ethics of care that involves designing around the concerns and needs of marginalized stakeholders is especially important given the uneven power relations between researchers and those stakeholders [39]. For instance, in discussing the data sources used, AIED researchers could reflect on their positionality concerning the participants and end users. Beyond informed consent as a standard practice in AIED, this includes acknowledging the additional roles and responsibilities implied for those providing the data to ensure they are not merely reduced to data subjects without rights and agency [7, 12]. For example, while surveillance and monitoring mechanisms in learning analytics may serve learners, they may also increase their vulnerability and raise concerns that could make learners less likely to benefit from AIED systems if left unexamined [12, 39]. Further, rather than building tools around feasibility, reflexivity creates a space to be transparent about dynamics during the design process of AIED tools, including how much weight was given to different stakeholders in design decisions. Transparency can then surface more ways in which AIED could incorporate critical theories and justice-driven design into its practices, supported by research community discussion.

In practice, achieving the level of transparency and reflexivity advocated may take much work for AIED researchers. Next to a lack of training resources for researchers or community platforms to engage in reflexivity, it is an open question how AIED researchers should best respond to discrepancies between current research goals and reflexivity that might question them. As research programs operate on medium-term time horizons of a few to several years, reflexivity could not only focus on broader research goals but also on smaller changes in research practices, such as prioritizing working with diverse samples, taking more time to solicit broad feedback from stakeholders during the AIED tool design process, or dedicating more time for needs finding rather than prioritizing rolling out novel capabilities early (e.g., generative AI-based AIED tools). This raises the question: Is the responsibility for ethical AIED research primarily at the individual researcher level, or does it necessitate broader institutional or community-wide commitment? Institutional and community-wide practices could play a critical role in creating the environment and providing the resources necessary for such ethical considerations to be integrated into the research process. To further encourage these practices, the AIED research community could advocate for including positionality statements and ethics requirements as part of the evaluation criteria for paper submissions. For instance, Cambo and Gergle [40] presented concepts of model positionality and computational reflexivity to encourage data scientists to reflect on the sociocultural contexts of model development, along with the backgrounds of annotators and researchers and their position within power dynamics with research subjects.

Establishing norms for conducting discussions that prioritize constructive

and critical engagement, alongside fostering a culture that values and rewards reflexivity in research activities, are essential steps toward establishing a more justice-oriented AIED field. Furthermore, AIED researchers could engage with wider education policies. For instance, to what extent is centering AIED research around institutional and national curricula or policies of standardized knowledge and skills desirable or constraining? How is AIED research hindered or enabled by wider cultural or policy factors, and to what extent might AIED researchers be positioned to challenge them? AIED researchers—along with policymakers, educators, learners, and other relevant stakeholders—could actively reflect on their positionalities and ask which constituencies will or will not benefit from the development and deployment of AIED systems: Whose perspectives are we looking from? Who benefits from such perspectives and is at a structural disadvantage [34]? Deliberately engaging with reflexivity kickstarts further initiatives to incorporate diverse voices from the wider community, creating more equitable and responsible AIED systems and practices.

4.2 Path 2: Increasing Diverse Stakeholder Collaboration and Advancing Research Methods for Diverse Stakeholder Involvement

As one solution of integrating concerns of AIED ethics into ethical AIED research practice, we have argued for including diverse perspectives and experiences in the design, development, and deployment of AIED systems. How can AIED develop, refine, and promote design research practices that listen to stakeholder needs and voices? Expanding methodologies for diverse stakeholder collaboration to envision desirable futures can involve integrating co-design principles and proactive adjustments to system design from the outset. Co-design methodologies emphasize the involvement of various stakeholders throughout the design process, ensuring that their perspectives, needs, and aspirations are incorporated into the final product or outcome [21]. This approach fosters inclusivity and ensures that the resulting solutions are more reflective of the diverse range of voices and experiences involved. Additionally, integrating changes to system design upfront allows for identifying and mitigating potential ethical concerns or unintended consequences early in the development process [41]. By actively involving stakeholders and considering ethical implications from the beginning, this approach promotes creating more robust, inclusive, and responsive solutions, fostering trust.

What does a coupling of inclusive co-design with continuous measurement look like in research practice? Related human-computer interaction methodologies emphasize aligned values with AIED stakeholders throughout the design and deployment lifecycle. While research methodologies such as community-based participatory research [27] have emphasized the inclusion of stakeholders in the *design* process of AIED systems, these collaborative efforts must be embedded in and extended through the *adoption* process of AIED systems in specific educational settings to ensure sustainable innovation [42]. Design-based implementation research methodology centers the design and implementation of educational tools around identifying "persistent problems of practice from multiple stakeholders' perspectives" from the very outset and is committed to "developing capacity for sustaining change in systems" [42, p. 142-243]. This is especially important for AIED as discrepancies might exist between what role AI should play in issues according to different

stakeholders [36]. AIED research can also take inspiration from participatory research models, such as Research-Practice-Industry Partnerships, aligning the design of AIED systems to practitioners' needs while incorporating a critical research lens [44].

4.3 Path 3: Combining Co-Design with Quantitative Measurement

One lesson learned from studying potential synergies between AIED ethics and ethical AIED research practice is that measurement-based approaches to strengthening ethics in AIED research are not necessarily bad but rather *limited*. Prioritizing measurement-based approaches is unlikely to eliminate all ethical issues in AIED systems (e.g., remediating historical underrepresentation of certain demographic groups in AIED system design). We propose that co-design with underrepresented stakeholders could be combined with regular measurements of variation in learning rates and other outcomes of interest in different learner populations to achieve ethical AIED research goals. Further, coupling inclusive co-design practices with quantitative measurement of educational effectiveness could derive more general principles that make AIED systems effective for different learner populations by comparing different design variations of systems.

A research opportunity in AIED exists to study whether community-based design through critical lenses can create more favorable learning outcomes (as measured through established measures of AIED learning environment effectiveness, such as learning gains and learning rates). Inclusive design could facilitate appropriate and sustainable adoption by creating intentional feedback loops that elevate the voices and needs of all involved parties, achieving desirable outcomes for practitioners and learners. Rather than being perceived as a constraint to innovation in AIED, justice-oriented ethics could leverage hidden design opportunities by paying attention to the "weak signals" in social and education systems. Within these systems, individuals from marginalized standpoints are best equipped to identify alternative solutions to systemic flaws, as they are more vulnerable to current risks and cognizant of the fundamental social regularities often invisible to those in more privileged or dominant positions [35]. As such, beyond addressing measurement-oriented improvements to AIED systems, AIED research could simultaneously serve as a justice-oriented dialogic space that proactively bridges the concerns and visions of different stakeholders involved and affected by AIED.

5 Summary and Outlook

The present study discussed the relationship between AIED ethics and ethical AIED, highlighting a gap between critical ethical perspectives on the AIED research goals and the practical methodologies to address ethical concerns. This discussion contributed paths to bridging both lenses. Researcher reflexivity regarding their standpoints and definitions of potential issues in AIED systems (e.g., bias and representation) offer one entry point to bridge the ethical frameworks of AIED with its research practice. Promising avenues for resulting research include advancing methods for co-design with marginalized communities, which could be combined with established learning measurements (e.g., learning rates) in relation to technology design. Further, studying the disciplinary overlap between AIED ethics and ethical

AIED, including systematic review papers and quantitative inquiry into topic centers, could guide synergy. Prioritizing a synthesis between AIED ethics and ethical AIED could make our research community more resilient in the face of rapidly advancing technology and AI, threatening public interest and trust in AIED systems. Acknowledging that paths toward ethical AIED are intricate and multifaceted, we hope this discussion fosters ongoing dialogue, collaboration, and reflexivity among researchers, practitioners, and the communities they aim to serve.

Acknowledgments. We extend our heartfelt thanks to the Learnest Ethical AI in Education Fellowship for providing the space and resources that enabled us, the authors, to collaborate.

References

1. Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S.B., Santos, O.C., Rodrigo, M.T., Cukurova, M., Bittencourt, I.I., Koedinger, K.R.: Ethics of AI in Education: Towards a Community-Wide Framework. Int. J. Artif. Intell. Educ. 32, 504–526 (2022).
2. Munn, L.: The uselessness of AI ethics. AI Ethics. 3, 869–877 (2023).
3. Mökander, J., Schuett, J., Kirk, H.R., Floridi, L.: Auditing large language models: a three-layered approach. AI Ethics. (2023).
4. Eubanks, V.: Digital Dead End: Fighting for Social Justice in the Information Age. The MIT Press (2011)
5. Baker, R.S., Hawn, A.: Algorithmic Bias in Education. Int. J. Artif. Intell. Educ. 32, 1052–1092 (2022).
6. Moats, D., Seaver, N.: "You Social Scientists Love Mind Games": Experimenting in the "divide" between data science and critical algorithm studies. Big Data Soc. 6 (2019)
7. Fox, A.: Educational research and AIED: Identifying ethical challenges. In: Holmes, W. and Porayska-Pomsta, K. (eds.) The Ethics of Artificial Intelligence in Education. Routledge (2022)
8. Holmes, W., Persson, J., Chounta, I.A., Wasson, B., Dimitrova, V.: Artificial Intelligence and Education a Critical View Through the Lens of Human Rights, Democracy and the Rule of Law. The Council of Europe, France (2022)
9. Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C. M.: Internet Research: Ethical Guidelines 3.0. Assoc. Internet Res. (2020)
10. Biesta, G.: Good education in an age of measurement: ethics, politics, democracy. Paradigm Publishers, Boulder, Colo (2010)
11. Holmes, W.: The Unintended Consequences of Artificial Intelligence and Education. Education International: Brussels, Belgium (2023)
12. Lee, H. H., Gargroetzi, E.: "It's like a double-edged sword": Mentor Perspectives on Ethics and Responsibility in a Learning Analytics–Supported Virtual Mentoring Program. J. Learn. Anal. 10, 85–100 (2023)
13. Williamson, B.: Datafication of Education: A Critical Approach to Emerging Analytics Technologies and Practices. In: Rethinking Pedagogy for a Digital Age. Routledge (2019)

14. Selwyn, N.: Distrusting educational technology: critical questions for changing times. Routledge, Taylor & Francis Group, New York ; London (2014)
15. Morgan, B., Hogan, M., Hampton, A.J., Lippert, A., Graesser, A.C.: The Need for Personalized Learning and the Potential of Intelligent Tutoring Systems. In: Handbook of Learning from Multiple Representations and Perspectives. Routledge (2020)
16. Hill, J.R., Hannafin, M.J.: Teaching and learning in digital environments: The resurgence of resource-based learning. Educ. Technol. Res. Dev. 49, 37–52 (2001)
17. Reich, J.: Failure to disrupt: why technology alone can't transform education. Harvard University Press, Cambridge, Massachusetts (2020)
18. Koedinger, K.R., Carvalho, P.F., Liu, R., McLaughlin, E.A.: An astonishing regularity in student learning rate. Proc. Natl. Acad. Sci. 120 (2023)
19. Karumbaiah, S., Ocumpaugh, J., Baker, R.S.: Context Matters: Differing Implications of Motivation and Help-Seeking in Educational Technology. Int. J. Artif. Intell. Educ. 32, 685–724 (2022)
20. Mu, T., Jetten, A., Brunskill, E.: Towards Suggesting Actionable Interventions for Wheel-Spinning Students. International Educational Data Mining Society (2020).
21. Holstein, K., McLaren, B.M., Aleven, V.: Co-Designing a Real-Time Classroom Orchestration Tool to Support Teacher–AI Complementarity. J. Learn. Anal. 6, 27–52 (2019)
22. Ogan, A., Walker, E., Baker, R., Rodrigo, Ma.M.T., Soriano, J.C., Castro, M.J.: Towards Understanding How to Assess Help-Seeking Behavior Across Cultures. Int. J. Artif. Intell. Educ. 25, 229–248 (2015)
23. Kwon, C., Butler, R., Stamper, J., Ogan, A., Forcier, A., Fitzgerald, E., Wambuzi, S.: Learning Analytics for Last Mile Students in Africa. In: Proceedings of the 13th Learning Analytics and Knowledge Conference. (2023)
24. Banks, J.A.: The Routledge international companion to multicultural education. Routledge, Taylor & Francis Group, New York ; London (2009)
25. Hancock, A. M.: When Multiplication Doesn't Equal Quick Addition: Examining Intersectionality as a Research Paradigm. Perspect. Polit. 5, 63–79 (2007)
26. D'Ignazio, C., Klein, L.F.: Data feminism. The MIT Press, Cambridge, Massachusetts (2020)
27. Harrington, C., Erete, S., Piper, A.M.: Deconstructing Community-Based Collaborative Design: Towards More Equitable Participatory Design Engagements. Proc. ACM Hum.-Comput. Interact. 3, 216:1-216:25 (2019).
28. Brossi, L., Castillo, A.M., Cortesi, S.: Student Centered Requirements for the Ethics of AI in Education. In: Holmes, W. and Porayska-Pomsta, K. (eds.) The Ethics of Artificial Intelligence in Education: Practices, Challenges, and Debates. Routledge, New York. 91-112 (2022)
29. Kizilcec, R.F., Lee, H.: Algorithmic Fairness in Education. In: Holmes, W. and Porayska-Pomsta, K. (eds.) The Ethics of Artificial Intelligence in Education: Practices, Challenges, and Debates. Routledge, New York. 174-202 (2022)
30. Zambrano, A.F., Zhang, J., Baker, R.S.: Investigating Algorithmic Bias on Bayesian Knowledge Tracing and Carelessness Detectors. In: Proceedings of

the 14th Learning Analytics and Knowledge Conference. 349–359 (2024)
31. Bird, K.A., Castleman, B.L., Song, Y.: Are algorithms biased in education? Exploring racial bias in predicting community college student success. J. Policy Anal. (2024)
32. Canto, M.: Global Information Society Watch 2019 Artificial intelligence: Human rights, social justice and development: Brazil. Instituto de Pesquisa em Direito e Tecnologia do Recife (2019)
33. Browne, J.: AI and Structural Injustice: A Feminist Perspective. In: Browne, J., Cave, S., Drage, E., and McInerney, K. (eds.) Feminist AI: Critical Perspectives on Algorithms, Data, and Intelligent Machines (2023)
34. Draude, C., Klumbyte, G., Lücking, P., Treusch, P.: Situated algorithms: a sociotechnical systemic approach to bias. Online Inf. Rev. 44, 325–342 (2019)
35. Treviranus, J.: Learning to learn differently. In: Holmes, W. and Porayska-Pomsta, K. (eds.) The ethics of artificial intelligence in education: practices, challenges, and debates. Routledge, Taylor & Francis Group, New York, NY (2023)
36. Rahm, L., Rahm-Skågeby, J.: Imaginaries and problematisations: A heuristic lens in the age of artificial intelligence in education. Br. J. Educ. Technol. 54, 1147–1159 (2023)
37. Prinsloo, P., Slade, S.: Big data, higher education and learning analytics: Beyond justice, towards an ethics of care. Big data and learning analytics in higher education: Current theory and practice, 109-124 (2017)
38. Ocumpaugh, J., Roscoe, R.D., Baker, R.S., Hutt, S., Aguilar, S.J.: Toward Asset-based Instruction and Assessment in Artificial Intelligence in Education. Int. J. Artif. Intell. Educ. (2024)
39. Prinsloo, P., Slade, S.: Student Vulnerability, Agency and Learning Analytics: An Exploration. J. Learn. Anal. 3, (2016)
40. Cambo, S.A., Gergle, D.: Model Positionality and Computational Reflexivity: Promoting Reflexivity in Data Science. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 1–19. Association for Computing Machinery, New York, NY, USA (2022)
41. Bhimdiwala, A., Neri, R.C., Gomez, L.M.: Advancing the Design and Implementation of Artificial Intelligence in Education through Continuous Improvement. Int. J. Artif. Intell. Educ. 32, 756–782 (2022)
42. Underwood, S.M., Kararo, A.T.: Design-Based Implementation Research (DBIR): An Approach to Propagate a Transformed General Chemistry Curriculum across Multiple Institutions. J. Chem. Educ. 98, 3643–3655 (2021)
43. Fishman, B.J., Penuel, W.R., Allen, A.R., Cheng, B.H., Sabelli, N.: Design-Based Implementation Research: An Emerging Model for Transforming the Relationship of Research and Practice. Teach. Coll. Rec. 115, 136–156 (2013)
44. Pautz Stephenson, S., Banks, R., Coenraad, M.: Outcomes of Increased Practitioner Engagement in Edtech Development: How Strong, Sustainable Research-Practice-Industry Partnerships will Build a Better Edtech Future. Digital Promise (2022)