Fair4AIED 2025: First International Workshop on Fairness in Algorithmic Decision-Making for Education

Clara Belitz[1 [0009−0002−1960−0914]], Nigel Bosch[1 [0000−0003−2736−2899]], Lan Jiang[1 [0009−0004−2764−0697]], Ioana Jivet[2 [0000−0002−8715−2642]], Badmavasan Kirouchenassamy[3 [0009−0003−6502−154X]], HaeJin Lee[1 [0009−0000−0260−0462]], Barbara Leporini[4 [0000−0003−2469−9648]], Francesca Maridina Malloci[5 [0000−0003−3287−4450]], Mirko Marras[5 [0000−0003−1989−6057]], Daniela Rotelli[3 [0000−0002−0943−6922]], and Frank Stinar[1 [0009−0000−3152−6281]]

[1] University of Illinois Urbana–Champaign, USA
{cbelitz2,pnb,lanj3,haejin2,fstinar2}@illinois.edu
[2] CATALPA, FernUniversität in Hagen, Germany
ioana.jivet@fernuni-hagen.de
[3] Sorbonne University, Paris, France
{badmavasan.kirouchenassamy, daniela.rotelli}@lip6.fr
[4] University of Pisa & ISTI-CNR, Pisa, Italy
barbara.leporini@unipi.it
[5] University of Cagliari, Cagliari, Italy
{francescam.malloci, mirko.marras}@unica.it

Abstract. The applications of artificial intelligence in education (AIED) are increasingly subject to scrutiny concerning the fairness of their automated decision-making processes (e.g., auto-grading, problem sequencing, personalization, interventions). However, little attention has been paid to the fairness of the procedures by which these algorithmic decisions are made, i.e., their procedural fairness. This procedural fairness is of critical importance; even seemingly fair decisions can be objectionable to students, teachers, and other stakeholders if the decision-making process itself is unfair. Without a structured framework, AIED applications risk perpetuating biases that disadvantage certain groups, deepen educational inequalities, and marginalize minority voices. Our full-day `Fair4AIED` workshop will take the first steps toward developing a blueprint to systematically integrate algorithmic fairness into AIED research, practice, and adoption. Participants are engaged in short-lightning talks, Q&A sessions, and discussions to explore opportunities to improve fairness and raise awareness of its importance for AIED. Subsequently, small-group discussions followed by a world-café-based session foster a structured, hands-on approach to envision the blueprint's content.

Keywords: Bias · Fairness · Procedural Fairness · Fair AI · Artificial Intelligence in Education · AI Ethics in Education.

1 Scope and Relevance

The growing adoption of AI in education is reshaping teaching and learning [22, 8, 7]. AI is now used in adaptive learning platforms [10], intelligent tutoring systems [18], automated grading [19], and personalization [3], as examples, enhancing education via tailored experiences, efficiency, and process improvement.

Despite these advancements, ensuring fair outcomes in AI for education remains a critical challenge from both research [2, 12, 4, 6] and regulatory[6] perspectives, due to the growing reliance on data-driven decision-making and the vast diversity of learners [9, 16, 14, 21]. For instance, AI systems can amplify biases from training data, leading to unfair outcomes, especially for marginalized groups [5, 23, 1]. Similarly, AI-powered essay grading may inherit subjective biases from human evaluators [20]. Moreover, the rapid emergence of generative AI applications in education, including decision-making systems that dynamically suggest content and produce feedback [15, 17] introduces additional fairness, ethics, and responsibility challenges that remain largely underexplored in current research.

Many studies prioritize overall accuracy and learning outcomes but neglect their impact on varied groups, leaving fairness an open question [22]. When fairness is considered, it is often based on arbitrary decisions or generic technical definitions that lack clear contextualization to education [1, 11]. Broadly, questions of fairness often focus on decision makers (e.g., teachers, course designers, researchers), rather than decision recipients (e.g., students) [13].

To bridge these gaps, a community-driven effort is needed to systematize fairness integration into AIED research, practice, and adoption, extending also to emerging domains such as generative AI and AI-driven decision-making systems. The `Fair4AIED` workshop fosters critical discussions on building an operational blueprint to achieve this goal. It engages participants in addressing key themes, including defining fairness, identifying at-risk groups, translating fairness into measurable metrics, mitigating bias, reporting fairness analysis into research, and auditing AIED systems for fairness. By bringing together the community, the workshop aims to set the foundation for establishing best practices to ensure that fairness is consistently and effectively considered in AIED.

The `Fair4AIED` workshop is highly relevant to AIED 2025, as the related community plays a key role in developing intelligent educational tools. Ensuring this community's perspectives on fairness are considered is crucial for a shared direction. Fully aligned with the conference's focus, the workshop fosters collaborative dialogue, contributing to a shared operational framework and strengthening a broader community of practice to systematize fairness in AIED.

2 Organization and Outcomes

The `Fair4AIED` workshop brings together researchers, practitioners, policymakers, and industry professionals whose expertise is essential for embedding fairness in AIED systems. In particular, researchers play a key role in designing

[6] As an example, EU Artificial Intelligence Act (https://artificialintelligenceact.eu/).

Table 1. Summary of Workshop Organization.

Session	Description
Morning Part	
Keynote Talk	A presentation highlighting advances and emerging challenges concerning fairness in AIED.
Lightning Talks	Short slide-based presentations delivered by participants, based on their submitted abstracts.
Abstracts' Discussion	Attendees engage in direct dialogue with presenters to discuss the abstracts' content.
Morning Summary	A summary sheet highlighting the key points from the keynote and lightning talks.
Afternoon Part	
Opening Briefing	Introduction to the afternoon objectives, structure, and activities, including the use of Slido.
Grounding Round(s)	Participants are divided into small groups to analyse thought-provoking prompts.
Grounding Report(s)	Each group presents key insights to the broader audience, highlighting multiple viewpoints.
World Café Interaction	A structured, rotating discussion format where each table focuses on a section of the blueprint.
Closing	Table hosts summarize final insights, followed by a closing session led by the chairs.

such systems, making them central to developing fairness-aware methodologies. Ethics and policy specialists shape the regulatory and governance frameworks that define how fairness is implemented, ensuring alignment with societal and legal standards. Industry professionals and edtech developers are responsible for deploying these systems in the real world, where fairness affects individuals.

The workshop is structured into sessions, distributed across the morning and the afternoon (Table 1). The morning part consists of two sessions. In the first session, we invite a keynote speaker to share the current research challenges related to the specific themes of the workshop (e.g., around students' and instructors' perceptions of procedural fairness; metrics, evaluations, and definitions of procedural fairness in education; effect of procedural fairness in education; key challenges and limitations such as abusing procedural fairness, control of stakeholder voice, and belongings) and potential future directions, followed by Q&A. In the second session, we will have short lightning talks from participants who had the opportunity to submit an abstract in advance of the workshop. Each presenter will have approximately 5 minutes to present, followed by a dedicated Q&A session. After the talks, each presenter will be seated at a table and other participants will be able to walk around and discuss with presenters about their specific abstracts. At the end of morning part, we offer a brief summary of key issues and works from the keynote and the abstracts presented, which will also

serve as a take-home flier to encourage workshop participants to reach out to the organizers and among each other as well as engage in the afternoon part.

In the afternoon part, the workshop includes small-group discussion sessions followed by a world-café-based session to foster a structured, hands-on approach towards the blueprint's drafting. To this end, the organizers first introduce themes and content, significance and relevance, objectives and outcomes, the format and activities planned for the afternoon. `Slido` is used to gather information about the participants and form diverse groups of 4 to 10 participants (resulting in 4–6 groups depending on attendance to the workshop). This introduction ends with thought-provoking prompts that encourage exploration (e.g., "What should an operational framework for fairness in AI for education include, from an interdisciplinary perspective?"). The workshop then covers two sessions. In the first session, participants are divided into small working groups, each tasked with analysing one of the aforementioned prompts from technical, ethical, and practical perspectives. Each group leverages diverse expertise and documents key takeaways — including challenges, solutions, and open questions — after a focused discussion led by a self-assigned moderator. Each group finally shares its key insights with the broader audience, summarizing main points and highlighting diverse perspectives, e.g., using a shared document/slide deck/whiteboard. Additional rounds may be conducted, time permitting, with each group focusing on a different prompt in each round. Then, in the second session, the workshop organizers introduce the World Café methodology, outlining its rules, goals, and structure, aiming to synthesize the insights from the rest of the workshop into an actionable blueprint. Each table, moderated by a host, focuses on a key blueprint section, ensuring guided discussions and continuity as participants rotate. Participants rotate between tables during different rounds: each round includes a recap by the table host, an open discussion, and a last part that aims to draft documentation capturing key takeaways and actionable steps. The workshop will end with each table host summarizing insights, followed by a final part led by the workshop chairs to refine the blueprint and explore post-workshop initiatives.

3 Organizers' Biography

Clara Belitz is a Doctoral Candidate at the University of Illinois Urbana-Champaign. Her research focuses on algorithmic fairness for intelligent tutoring systems, with an emphasis on student voices and experiences. She also researches surveillance technologies and the appropriate uses of data for personalization. She is also a community organizer.

Nigel Bosch is an Assistant Professor at the University of Illinois Urbana–Champaign, with a joint appointment in Information Sciences and Educational Psychology. His research concerns machine learning, algorithmic fairness, human-centered computing, notably in educational contexts.

Lan Jiang is a Doctoral Student at the University of Illinois Urbana–Champaign. Her current research focuses on developing automated text assessment systems across various contexts and on identifying and reducing bias within these

systems. Her broader research interests include understanding student learning behaviors and enhancing learning experiences through natural language processing and machine learning.

Ioana Jivet is a Junior Research Professor leading the Learning Analytics group at CATALPA, FernUniversität in Hagen. Her main research interests focus on providing feedback in higher education via student-facing learning analytics, including designing learning analytics dashboards and generating textual feedback. She is also researching cultural factors that influence the personalization and uptake of learning analytics among students and teachers. She has a MSc in Computer Science from TU Delft and a PhD in Learning Analytics from the Open University of the Netherlands.

HaeJin Lee is a Doctoral Student at University of Illinois Urbana–Champaign. Her research lies at the intersection of human-computer interaction, learning analytics, and explainable AI in educational contexts. Specifically, she focuses on developing and evaluating learning interventions using explainable AI methods in computer-based learning environments.

Badmavasan Kirouchenassamy is a Doctoral Student at Sorbonne University, specializing in the development of automated adaptive feedback systems for online programming education. His research focuses on leveraging reinforcement learning techniques to enhance real-time decision-making. His interests extend also to related areas, including automated error detection in student-written code and feedback generation using information retrieval techniques. His work aims to improve the efficiency and personalization of learning experiences in programming education.

Barbara Leporini is an Associate Professor of Computer Science at the University of Pisa, previously a senior researcher at ISTI-CNR. Her research focuses on Human-Computer Interaction, particularly methods and tools for accessible and inclusive digital interfaces. She has led and participated in numerous national and international projects on accessibility, assistive technologies, and education. She is actively engaged in educational and accessibility policy, contributing to defining accessibility requirements in Italian law. She currently directs a training agency and serves on the Board of Directors of the National Institute for Research, Training, and Rehabilitation (I.Ri.Fo.R.).

Francesca Maridina Malloci is an Assistant Professor at the University of Cagliari, specializing in predictive analytics and recommender systems, with a strong focus on decision-making algorithms in multi-stakeholder contexts, including education. She has published extensively in international journals and conference proceedings, contributing to the intersection of AI, fairness, and education. Her current research explores how recommender systems can support fair and transparent decision-making in educational settings. In 2019, she was a visiting scientist at the EURECAT Technology Centre (Spain), collaborating with the Data Science and Big Data Analytics Unit.

Mirko Marras is a Tenure-Track Assistant Professor at the University of Cagliari, specializing in responsible AI, with a strong focus on AI in education and fairness. He has published over 80 papers in top conferences and journals and

serves on program committees for major venues, including AIED, EDM, LAK, ICALT, EC-TEL, and UMAP. He has co-chaired multiple workshops, including BIAS (ECIR 2020-2023), FATED (EDM 2022), L2D (WSDM 2021), and RKDE (ECML-PKDD 2023-24). He has been General Co-Chair of ACM UMAP 2024, Blue Sky Track Co-Chair at AIED 2024, Proceedings Co-Chair at EDM 2023-25, and IR for Good Track Co-Chair at ECIR 2024 and 2025. He was a Postdoc at EPFL in 2021.

Daniela Rotelli is a Postdoctoral Researcher at Sorbonne University (France), specialising in learning analytics and educational data mining. She earned her doctoral degree in Computer Science at the University of Pisa (Italy), focussing on data preprocessing and modelling student behaviour in online learning environments. Her research aims to develop data-driven frameworks for understanding temporal learning behaviours, identifying trends, predicting challenges, and supporting fair and effective interventions.

Frank Stinar is a Doctoral Student at the University of Illinois Urbana–Champaign. His research focuses on how normative ethics, algorithmic fairness, and student modeling all interact. Broadly, his research aims to make education more equitable.

References

1. Baker, R.S., Hawn, A.: Algorithmic bias in education. International journal of artificial intelligence in education pp. 1–41 (2022)
2. Balloccu, G., Boratto, L., Cancedda, C., Fenu, G., Marras, M.: Knowledge is power, understanding is impact: Utility and beyond goals, explanation quality, and fairness in path reasoning recommendation. In: European Conference on Information Retrieval. pp. 3–19. Springer (2023)
3. Bernacki, M.L., Greene, M.J., Lobczowski, N.G.: A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose (s)? Educational Psychology Review **33**(4), 1675–1715 (2021)
4. Boratto, L., Fenu, G., Marras, M., Medda, G.: Practical perspectives of consumer fairness in recommendation. Information Processing & Management **60**(2), 103208 (2023)
5. Burriss, S.K.: An AI ethics primer for educators. Revolutionizing English Education: The Power of AI in the Classroom p. 179 (2024)
6. Cock, J.M., Bilal, M., Davis, R., Marras, M., Kaser, T.: Protected attributes tell us who, behavior tells us how: A comparison of demographic and behavioral oversampling for fair student success modeling. In: LAK23. pp. 488–498 (2023)
7. Dessí, D., Dragoni, M., Fenu, G., Marras, M., Reforgiato Recupero, D.: Deep learning adaptation with word embeddings for sentiment analysis on online course reviews. Deep learning-based approaches for sentiment analysis pp. 57–83 (2020)
8. Dessì, D., Fenu, G., Marras, M., Reforgiato Recupero, D.: Leveraging cognitive computing for multi-class classification of e-learning videos. In: The Semantic Web: ESWC 2017 Satellite Events: ESWC 2017 Satellite Events, Portorož, Slovenia, May 28–June 1, 2017, Revised Selected Papers 14. pp. 21–25. Springer (2017)
9. Drachsler, H., Hoel, T., Scheffel, M., Kismihók, G., Berg, A., Ferguson, R., Chen, W., Cooper, A., Manderveld, J.: Ethical and privacy issues in the application

of learning analytics. In: Proceedings of the Fifth International Conference on Learning Analytics And Knowledge. pp. 390–391. ACM, Poughkeepsie New York (Mar 2015). https://doi.org/10.1145/2723576.2723642
10. Dutta, S., Ranjan, S., Mishra, S., Sharma, V., Hewage, P., Iwendi, C.: Enhancing educational adaptability: A review and analysis of ai-driven adaptive learning platforms. In: 2024 ICIPTM. pp. 1–5. IEEE (2024)
11. Fenu, G., Galici, R., Marras, M.: Experts' view on challenges and needs for fairness in artificial intelligence for education. In: AIED. pp. 243–255. Springer (2022)
12. Gómez, E., Shui Zhang, C., Boratto, L., Salamó, M., Marras, M.: The winner takes it all: geographic imbalance and provider (un) fairness in educational recommender systems. In: ACM SIGIR. pp. 1808–1812 (2021)
13. Heuer, L.: What's just about the criminal justice system - A psychological perspective. Journal of Law and Policy **13**(1), 209–228 (2005), https://brooklynworks.brooklaw.edu/jlp/vol13/iss1/11
14. 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. International Journal of Artificial Intelligence in Education **32**(3), 504–526 (Sep 2022). https://doi.org/10.1007/s40593-021-00239-1
15. Hu, X., Xu, S., Tong, R., Graesser, A.: Generative AI in education: From foundational insights to the socratic playground for learning. arXiv preprint arXiv:2501.06682 (2025)
16. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nature Machine Intelligence **1**(9), 389–399 (Sep 2019). https://doi.org/10.1038/s42256-019-0088-2
17. Lindsay, E.D., Zhang, M., Johri, A., Bjerva, J.: The responsible development of automated student feedback with generative ai. arXiv:2308.15334 (2023)
18. Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., Ghazi Saeedi, M.: Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. ILE (2021)
19. Ramesh, D., Sanampudi, S.K.: An automated essay scoring systems: a systematic literature review. Artificial Intelligence Review **55**(3), 2495–2527 (2022)
20. Review, T.P.: The evolution of education: How ai is reshaping grading (2024), https://www.princetonreview.com/ai-education/how-ai-is-reshaping-grading, accessed: March 2, 2025
21. Slade, S., Prinsloo, P.: Learning Analytics: Ethical Issues and Dilemmas. American Behavioral Scientist **57**(10), 1510–1529 (Oct 2013). https://doi.org/10.1177/0002764213479366
22. Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., Du, Z.: AI in education: A systematic literature review. Expert Systems with Applications (2024)
23. Yu, R., Li, Q., Fischer, C., Doroudi, S., Xu, D.: Towards accurate and fair prediction of college success: Evaluating different sources of student data. International educational data mining society (2020)