Toward Intersectionally Fair Educational AI for Learning Analytics and Student Guidance

Stefania Zourlidou, Tai Le Quy, Resmin Hossain and Frank Hopfgartner
University of Koblenz, Koblenz, Germany

zourlidou@uni-koblenz.de
taileguy@uni-koblenz.de
resminhossain@uni-koblenz.de
hopfgartner@uni-koblenz.de

Abstract: Artificial intelligence is increasingly embedded in learning analytics dashboards, adaptive learning environments and student-guidance tools. These systems do more than personalise learning: they rank, filter and prioritise educational opportunities in ways that shape learner visibility, access and future pathways. Fairness evaluations, however, often remain organised around single protected attributes such as gender, ethnicity, socio-economic status or disability. This paper argues that such evaluations are insufficient for educational contexts because they can obscure harms experienced by learners located at the intersection of multiple demographic positions. It therefore advances intersectional fairness auditing as a human-centred requirement for responsible educational AI. The argument is informed by a motivating adjacent-domain case: a study of a transformer-based job-recommendation system found that a system appearing fair attribute by attribute can still conceal substantial unfairness for intersectional subgroups, and that common mitigation methods do not necessarily improve all subgroups evenly. Drawing on recent scholarship on human-centred AI in education, learning analytics auditing, educational recommendation and intersectional fairness, the paper proposes an audit framework for educational AI. The framework extends existing learning analytics audit work by making subgroup visibility, ranking effects, demographic data governance, contestability and post-deployment monitoring explicit. The paper concludes that educational AI should be judged not only by predictive accuracy or personalisation, but also by whether it supports equitable learner futures across intersecting differences and enables learners and educators to question, interpret and contest AI-mediated guidance.

Keywords: Educational AI, Intersectional Fairness, Learning Analytics, Student Guidance, Human-Centred AI, Audit Framework

1. Introduction

Artificial intelligence is now woven into many educational decision-support environments: dashboards highlight learners at risk, recommenders suggest courses and pathways, admissions systems allocate scarce places, and career-guidance tools mediate how students pursue future opportunities. Although often discussed in terms of efficiency and personalisation, they also shape visibility, access and trajectories by ranking, filtering and prioritising some options over others. This makes fairness a central educational concern. Unfairness can influence whether a learner sees a recommendation, is flagged for support, or is channelled towards particular futures. Baker and Hawn (2022) argue that algorithmic bias in education is not simply a matter of isolated model defects, but a broader issue involving design choices, data practices and institutional contexts. Yet much fairness work in education still treats identity one axis at a time. Gender, ethnicity, socio-economic status or disability may each be analysed separately, but learners do not experience educational systems one attribute at a time. A mature student from a minority ethnic background, a disabled student from a low-income household, or a first-generation international student each occupies an intersectional position. Single-axis audits may therefore make an educational AI system appear fair while obscuring patterned harms for smaller or less visible subgroups.

This paper argues that educational AI needs an explicitly intersectional fairness perspective. It builds this argument through a conceptual, case-informed approach. The motivating case comes from a recent master's thesis on a transformer-based job recommendation system. That work found that a system appearing fair on single protected attributes could still generate intersectional unfairness, and that technical mitigation strategies did not improve outcomes evenly across subgroups (Hossain, 2026). This is not introduced as evidence about real hiring practice, but as an adjacent high-stakes recommender case from which education can learn. The paper makes four contributions: it synthesises recent literature on responsible AI in education, fairness in learning analytics and educational recommendation; reframes many educational AI systems as allocating opportunity rather than only personalising; transfers lessons from an intersectional job-recommendation case to learning analytics and student guidance; and proposes an intersectional fairness audit framework foregrounding subgroup visibility, ranking effects, contestability and institutional accountability. Based on the literature and the motivating case, the paper addresses three research questions: (RQ1) How do recent studies on educational AI, learning analytics and student-guidance systems conceptualise fairness, and to what extent

do they address intersectional harms? (RQ2) What transferable lessons from intersectional bias in job-recommendation systems are most relevant for learning analytics and student guidance in higher education? (RQ3) What should an intersectional fairness audit framework for educational AI include if it is to support human-centred and responsible deployment?

2. Recent Literature Review

Recent work on educational AI has moved beyond simple enthusiasm for automation towards stronger attention to human-centred design, stakeholder participation and hybrid intelligence. Alfredo et al. (2024), in a systematic review of human-centred learning analytics and AI in education, show that stakeholder involvement is increasingly recognised as necessary, but remains uneven across the design process. Their review connects safety, reliability and trustworthiness to when and how users are involved. Topali et al. (2025) reinforce this concern from a more design-focused angle. Reviewing 47 empirical studies on human-centred learning analytics and AI in education, they report that stakeholder participation is often present but concentrated in later stages, while ideation, prototyping and evaluation remain less consistently reported. Their absence during early design can normalise narrow assumptions about what counts as a good outcome or recommendation. Cukurova (2025) pushes the conversation further by framing AI in education through the lens of hybrid intelligence. By treating AI not purely as a tool that replaces educational judgement, he argues for a broader understanding in which AI systems extend, challenge and reshape human cognition and practice. Fairness therefore cannot be reduced to outputs alone, it also concerns how humans interpret, act on and contest those outputs. Hooshyar et al. (2025) identify nine persistent challenges in current AI for education and argue that inadequate stakeholder involvement and weak ethical oversight remain central weaknesses. Their proposal for hybrid human-AI systems resonates strongly with a fairness agenda. If educational AI is hybrid and not autonomous, then fairness must include human oversight, interpretability, contestability and governance, not just model optimisation.

Within learning analytics, auditing and governance are becoming more visible themes. Simbeck (2024) proposes a four-phase audit process for learning analytics systems consisting of delimitation, risk-based definition of audit criteria, auditing and assessment, and monitoring and reassurance. This provides an education-specific bridge between general AI ethics principles and the practical evaluation of deployed systems. Importantly, Simbeck centres the learner perspective and argues that learning analytics should be audited for whether they benefit learners and avoid placing them at risk. Baker and Hawn (2022) also remain foundational for educational fairness because they map algorithmic bias across the educational pipeline. Their analysis makes clear that bias can emerge from the framing of educational problems, the collection and labelling of data, the modelling stage, and institutional implementation. This pipeline view aligns closely with the argument of the motivating job-recommendation thesis we use in this study, where unfairness was treated as a property of the broader model-data interaction and not a simple reflection of a single model component. A more application-specific picture emerges in Cabral et al. (2025), whose systematic review of AI-powered learning analytics dashboards found that most systems focus on predicting academic performance, supporting self-regulated learning and providing teacher-facing insights. The review documents rapid application growth while also pointing to recurring research gaps, including limited evidence about classroom deployment and the long-term effects of dashboard use. Dashboards are not neutral windows onto student behaviour, but they structure attention and determine which learners appear visible, urgent or successful.

The relevance of fairness becomes clearer when educational AI is understood as recommendation and guidance infrastructure. Shailya et al. (2025) examine fairness in language-model-based academic recommendations and identify geographic, demographic and economic biases in university and programme suggestions. Their work is especially useful for this paper because it shows bias at the ranking and visibility level: some institutions, countries and learner profiles become more recommendable than others. Weissburg et al. (2025) reach a related conclusion from another angle. Evaluating large language models as personalised educational teachers, they show significant bias in the generation and selection of educational content across demographic groups, including race, ethnicity, sex, gender, disability, income and national origin. This is not a traditional recommender setting, but it shows that bias in educational AI extends beyond ranking systems to personalised explanation, support and feedback. Fairness concerns also reach AI-supported assessment: Zacharis and Papadakis (2025) examine the validity, reliability and fairness of AI-graded university coursework, indicating that responsible educational AI must meet the same fairness standards whether it ranks, generates or grades. Student guidance is also an emerging site of AI-mediated decision making. A systematic review by Sarmurzin et al. (2026) shows that AI is transforming university career counselling from predictive tools towards interactive generative agents, while simultaneously raising concerns about student agency and the platformisation of

guidance. In a related but different context, Ibrahim et al. (2025) present an equity-aware recommender for university admissions, showing how recommendation and allocation methods are increasingly implicated in high-stakes educational decisions. These studies suggest that educational AI increasingly allocates opportunity instead of merely offering information. When a system influences what learners see, which pathways appear plausible and who receives support or visibility, fairness becomes inseparable from educational design.

Intersectionality offers a language for harms that are not visible when systems analyse social categories separately. Crenshaw's foundational work (Crenshaw, 1989) showed how legal and institutional frameworks can marginalise people whose experiences sit at the intersection of multiple structures of disadvantage. In algorithmic systems, Buolamwini and Gebru (2018) provided an influential demonstration of this problem: in commercial gender-classification systems, darker-skinned females were the most misclassified group, with error rates of up to 34.7%, while the maximum error rate for lighter-skinned males was 0.8%. Their study is outside education, but its lesson is directly relevant, as apparently acceptable aggregate performance can conceal substantial harm for intersectional subgroups. Despite these foundational insights, intersectionality remains underdeveloped in educational AI research. Mirsch et al. (2025) explicitly argue that single-axis approaches to fairness in educational AI obscure the experiences of individuals facing intersecting forms of discrimination. Their intervention is important because it makes visible a recurring blind spot, showing that educational fairness conversations often stop at disaggregating by one variable at a time. The motivating work by Hossain (2026) provides a useful adjacent-domain illustration of why this matters. In a transformer-based job recommendation setting, the study found that fairness assessed on a single protected attribute could mask substantial unfairness across combined demographic subgroups, and that commonly used mitigation methods did not produce balanced improvement across all groups. Although the application domain differs from education, the structural similarity is strong. Both job recommendation and educational AI rank opportunities, define positive outcomes, and shape future trajectories under conditions of partial visibility and unequal data representation. Read across these strands, the reviewed studies engage several distinct conceptions of fairness, but unevenly. Performance fairness, the parity of predictive accuracy or error across groups, is the most established concern and underlies pipeline accounts of bias (Baker and Hawn, 2022) and demonstrations that aggregate accuracy hides subgroup disparity (Buolamwini and Gebru, 2018). Representational fairness, concerning who is visible or well served in data and generated content, appears in work on biased educational content (Weissburg et al., 2025). Exposure or ranking fairness, concerning which options are made visible in top-k results, is beginning to be examined in academic recommendation and allocation (Shailya et al., 2025; Ibrahim et al., 2025). Procedural fairness, namely auditing, oversight and reassurance, is articulated most clearly by Simbeck (2024) and Hooshyar et al. (2025), while participatory fairness, the early involvement of learners and educators, is foregrounded in human-centred and hybrid-intelligence accounts (Alfredo et al., 2024; Topali et al., 2025; Cukurova, 2025). By contrast, contestability, whether learners can question or challenge an AI-mediated outcome, is gestured at but rarely operationalised, and intersectional fairness remains thin, raised directly only by Mirsch et al. (2025). In answer to RQ1, recent educational-AI scholarship conceptualises fairness mainly through performance, participation and emerging audit practice, with exposure fairness newly visible, whereas intersectional harm and contestability are the least developed, which is precisely the gap the proposed framework addresses.

Table 1: Selected recent literature and its relevance to this paper

Study	Domain	Main Relevance to this Paper
Alfredo et al. (2024)	Human-centred LA/AI	Shows that user participation is important but uneven across design stages.
Topali et al. (2025)	Human-centred design	Finds stakeholder involvement is common but often late, with evaluation underreported.
Cukurova (2025)	Hybrid intelligence	Positions educational AI as part of human–AI judgement rather than replacement.
Simbeck (2024)	Learning analytics audit	Provides an education-specific audit model focused on learner risk and reassurance.
Cabral et al. (2025)	AI-powered dashboards	Shows expanding dashboard use in prediction, self-regulation and teacher insight.

Study	Domain	Main Relevance to this Paper
Shailya et al. (2025)	Academic recommendations	Demonstrates geographic, demographic and economic bias in educational recommendation rankings.
Weissburg et al. (2025)	Personalised education	Shows bias in LLM-generated educational content across several demographic groups.
Crenshaw (1989)	Intersectionality	Provides the foundational concept for analysing harms that arise at intersections of social categories.
Buolamwini and Gebru (2018)	Algorithmic fairness	Shows how aggregate system performance can conceal substantial disparities for intersectional subgroups.
Mirsch et al. (2025)	Intersectional fairness	Argues that single-axis fairness approaches obscure intersecting discrimination in educational AI.

3. Approach

This paper adopts a conceptual, case-informed design that combines a targeted integrative review of recent scholarship with reflective transfer from an adjacent-domain thesis case. It is not a full systematic review. Instead, we deliberately gathered the studies needed to make sense of intersectional fairness in educational AI, and we explain how we did so, so that readers can judge our choices for themselves. Sources were identified through Scopus, Web of Science, the ACM Digital Library, IEEE Xplore and Google Scholar, supplemented by hand-searching recent proceedings (ACL/NAACL, IJCNLP, PMLR) and the journals Computers and Education: Artificial Intelligence, British Journal of Educational Technology and AI and Ethics. Search terms combined fairness vocabulary (fairness, algorithmic bias, intersectional) with educational and system vocabulary (learning analytics, educational AI, student guidance, academic recommendation, audit). Studies were included if they were peer-reviewed, published between 2024 and 2026, written in English, and addressed at least one of four strands: human-centred and hybrid AI in education, fairness and auditing in learning analytics, educational recommendation and guidance, or intersectional fairness. Work outside these strands and non-peer-reviewed material were excluded, with the single, explicitly flagged exception of the motivating thesis, which is used illustratively. Pre-2024 work was excluded except for foundational sources on intersectionality and algorithmic bias (Crenshaw, 1989; Buolamwini and Gebru, 2018; Baker and Hawn, 2022), retained because they remain conceptually indispensable. The 2024 to 2026 window was chosen because generative and transformer-based models were rapidly integrated into educational AI during this period, and because the most relevant recent reviews and the motivating case are contemporaneous, allowing the transfer of lessons to be situated against current debates rather than an outdated baseline. This recency window applies only to evidence about how current systems behave. The framework's ethical and governance foundations draw on established learning-analytics-ethics scholarship that predates it and remains authoritative, including early analyses of the ethical dilemmas of learning analytics (Slade and Prinsloo, 2013) and checklist-based approaches to trusted, privacy-aware analytics (Drachsler and Greller, 2016). Selected studies were analysed thematically: for each, the domain, the system type and, central to RQ1, how it conceptualises and operationalises fairness were recorded. The motivating case is a master's thesis (Hossain, 2026), openly available under a Creative Commons licence and examined at the University of Koblenz, which audited a transformer-based job-recommendation pipeline. The system matched applicant resumes to O*NET occupational profiles using Sentence-BERT embeddings (all-mpnet-base-v2) and cosine similarity, evaluated as a top-k recommender. Fairness was assessed with three metrics, namely demographic parity, conditional demographic parity and equal opportunity, first for single attributes (gender, race) and then for intersectional subgroups combining gender, race, age and experience level, using a publicly available recruitment dataset of mixed synthetic and real application data. The study found that a system appearing fair attribute by attribute concealed substantial disparities for smaller intersectional subgroups, that two layers of resampling improved fairness unevenly and traded accuracy against parity, and that aggregate top-k fairness masked persistent ranking disadvantage for some groups. The thesis is used here as an illustrative, adjacent-domain case that motivates the framework, not as independent empirical evidence about educational systems. No fairness scores are reproduced, and no claim about education rests on it as if externally validated. The transfer from job recommendation to education is therefore analytical, not literal. Both domains use data-driven systems to rank, filter and narrate future opportunities under conditions of unequal representation, which is what makes the analogy useful. But the two differ, and those differences set limits.

Education carries its own institutional and legal duties, teaching relationships, developmental goals, and expectations about student choice, so the job-recommendation lessons are best treated as questions to ask about educational AI, not answers to assume.

4. From Job Recommendation to Educational AI

4.1 Educational AI as Opportunity Allocation

A central argument of this paper is that many educational AI systems should be understood as opportunity-allocation systems. This section addresses RQ2 by identifying the transferable lessons from intersectional bias in job-recommendation systems that are most relevant for learning analytics and student guidance. They allocate not scarce seats but visibility, salience, intervention and perceived possibility. In education, the consequences of ranking and filtering are normative. A learner may never encounter the course, placement, institution or support pathway that would have mattered most because the system ranked alternatives differently. In both domains, top-k exposure matters, outcome definitions matter, and apparently small ranking effects can have long-term trajectory consequences.

4.2 Why Single-Attribute Fairness is Insufficient

The second argument is that single-attribute fairness analysis is inadequate for educational AI. If a system is audited only for gender parity, it may still disadvantage older women, disabled women, or women from specific ethnic groups. If it is audited only for socio-economic status, it may still reproduce harms for low-income international students or mature learners from minority backgrounds. The point is not merely combinatorial complexity. It is that social disadvantage is often produced at intersections, not along one line at a time. Mirsch et al. (2025) make this point directly for educational AI. The study by Hossain (2026) illustrates how this arises within a recommendation pipeline: when fairness was examined attribute by attribute, the system seemed relatively fair, but when subgroup combinations were evaluated, substantial disparities became visible. This lesson translates well to learning analytics and student guidance because these systems also work with heterogeneous populations in which smaller subgroups are easily drowned out by majority patterns. Institutions are especially prone to missing this because they treat AI as support rather than selection infrastructure, yet support systems still stratify opportunity by shaping who is noticed and encouraged.

4.3 Why Technical Fixes Alone are Insufficient

A third lesson from the motivating case is that bias mitigation should not be treated as a one-step technical remedy. In Hossain (2026), preprocessing interventions improved some fairness outcomes but did not improve all subgroups evenly. Practitioners should therefore not deploy a mitigation method and assume the problem is solved. A mitigation technique can reduce disparity for a visible aggregate group while shifting error, exposure or risk onto a smaller intersectional subgroup. Hooshyar et al. (2025) emphasise inadequate ethical oversight and stakeholder involvement, while Simbeck (2024) argues for ongoing monitoring and reassurance and not just one-off checks. Together these works suggest that responsible educational AI requires socio-technical governance. Fairness must be revisited as systems are updated, populations change and uses drift.

5. An Intersectional Fairness Audit Framework for Educational AI

Drawing on the literature and the motivating case, this paper proposes an intersectional fairness audit framework for educational AI systems used in learning analytics, student guidance, academic recommendation and related decision-support settings. In doing so, the section responds to RQ3 by specifying what an intersectional fairness audit framework for educational AI should include to support human-centred and responsible deployment. The framework builds on Simbeck's (2024) learning analytics audit model and earlier trusted-analytics checklists (Drachsler and Greller, 2016), but it differs in three ways. First, it makes intersectional subgroup visibility an explicit audit object and not just treating fairness as a general risk category. Second, it is ranking-sensitive, recognising that many educational AI systems distribute visibility and exposure rather than only binary decisions. Third, it connects technical evaluation to human-centred practices of explanation, contestability and institutional monitoring. The framework is intended as a practical checklist for design, procurement, deployment and post-deployment review. It does not prescribe a single fairness metric, instead, it asks institutions to make explicit what the system does, who may be affected, what outcomes and rankings matter, and how learners can challenge or reinterpret AI-mediated guidance.

Table 2: Proposed intersectional fairness audit framework

Audit Dimension	Core Question	Lifecycle Point	Illustrative Practices
Context and stakes	What educational decision or guidance function is the system shaping?	Design and procurement	Define the educational use, affected learners and possible trajectory effects before evaluation.
Data representation and demographic governance	Who is visible in the data, and who may be missing, proxied or flattened by aggregation?	Design and data preparation	Review coverage, proxies, subgroup size, data minimisation, consent and small-cell re-identification risks.
Outcome and ranking definition	What counts as a positive outcome, and how does ranking influence exposure?	Design and evaluation	Audit top-k visibility, recommendation depth and which options are systematically foregrounded.
Intersectional evaluation	Do subgroup combinations reveal harms that single-attribute checks miss?	Testing and deployment	Compare aggregate results with ethically appropriate subgroup combinations and report uncertainty for small groups.
Transparency and explanation	Can learners and educators understand the system's purpose, limits and logic?	Deployment	Provide intelligible explanations, model cards or user-facing guidance on system scope and uncertainty.
Human oversight and contestability	Can outputs be reviewed, overridden or contested?	Deployment and use	Build teacher review processes, learner feedback channels and routes for appeal or correction.
Governance and monitoring	Who is responsible for fairness after deployment?	Post-deployment	Assign institutional ownership, document updates and run periodic audits across cohorts and contexts.

To show what the framework changes in practice, consider a recommender that suggests optional first-year modules, where some modules feed into competitive, employability-linked progression routes. A conventional single-axis audit would check overall click-through or completion and perhaps compare recommendations for men and women, conclude that parity holds, and stop. The framework reorganises this audit around educational consequence. Under context and stakes, the team first establishes that the system is not merely informational: by shaping which routes appear plausible, it allocates opportunity. Under data representation, it examines who is present in the training data and at what granularity, and here the motivating case is instructive, because the intersectional cells most exposed to harm, such as the smallest subgroup in Hossain (2026) with only a few dozen records, are exactly those a majority-driven model represents least reliably. This raises an ethical and statistical limit: only subgroup combinations large enough to support a defensible claim should be reported, small-cell counts must be disclosed and not silently pooled, and re-identification risk constrains how finely groups may be cut. Under outcome and ranking definition and intersectional evaluation, the audit inspects top-k exposure across combined groups rather than aggregate rates, since the case showed that aggregate fairness can improve while relative ranking still disadvantages particular subgroups, and that resampling can buy parity at the cost of accuracy, so both must be reported together with uncertainty for small groups. Finally, transparency, human oversight and contestability, and governance require that learners receive intelligible explanations that counter automation bias, that a teacher can override a steer and a learner can challenge it through a defined route, and that a named owner re-audits as cohorts change. A single-axis audit would have passed this system as fair. The intersectional audit instead reveals who stays invisible, where ranking quietly widens gaps in opportunity, and how learners can challenge the results.

6. Discussion

The proposed framework has practical implications for several educational contexts. Returning to RQ1, the literature reviewed in Section 2 conceptualises fairness in educational AI primarily through human-centred design, learning analytics auditing and bias in recommendation or generation, but the reviewed studies address intersectional harms only partially, with most contributions still organised around single protected attributes. The framework proposed here addresses that gap. Across learning analytics dashboards, student-guidance

systems and admissions tools, it shifts attention from overall model performance to whose behaviour is legible, whose needs become visible, how recommendations may narrow aspiration, and the distributive consequences of ranking and allocation.

This paper has limitations. First, it makes a conceptual contribution. It does not report new empirical findings, and the review covers a focused set of recent studies. The framework therefore needs empirical testing in specific educational settings such as learning analytics dashboards, academic guidance chatbots and institutional career services. Future research could operationalise the framework through case studies, co-design workshops, or audits of existing university systems. A further step is to examine how ethical and institutional constraints affect intersectional evaluation, especially where demographic data are incomplete or sensitive. The current educational AI conversation is rapidly expanding, but it risks reproducing a familiar pattern in which fairness language is adopted while single-axis evaluations and weak governance remain in place. Intersectional auditing provides a clearer basis for responsible practice because it asks institutions to identify who benefits, who remains invisible and who can contest the system's effects.

7. Conclusion

Educational AI increasingly ranks, filters and guides learners across key moments of academic and professional development. As these systems expand, fairness must be treated as an educational and institutional concern, not merely a technical feature. Recent literature shows growing awareness of bias and responsibility, but intersectionality remains insufficiently integrated into practice. In response, the paper proposed an intersectional fairness audit framework centred on context, data representation and governance, ranking, intersectional evaluation, transparency, human oversight and monitoring. Educational AI should be judged not only by whether it predicts well or personalises efficiently, but by whether it supports equitable participation across intersecting learner identities.

Acknowledgements

The authors received no specific funding for this work.

Ethics Declaration

This paper is a conceptual study drawing on published literature. It reports no new human-participant research and no new collection of personal data. The motivating case (Hossain, 2026) is a publicly available, independently examined master's thesis released under a Creative Commons licence. Only its published findings are used, and no non-public or author-held personal data were drawn upon. Ethical clearance was therefore not required for the research reported here.

AI Declaration

An AI language model was used solely for language revision. All substantive arguments, interpretations, source selections, and final wording were reviewed and approved by the authors, who take full responsibility for the final manuscript.

References

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D. and Martinez-Maldonado, R. (2024) "Human-centred learning analytics and AI in education: A systematic literature review", *Computers and Education: Artificial Intelligence*, 6, 100215. doi:10.1016/j.caeai.2024.100215.

Baker, R.S. and Hawn, A. (2022) "Algorithmic bias in education", *International Journal of Artificial Intelligence in Education*, 32, pp. 1052–1092. doi:10.1007/s40593-021-00285-9.

Buolamwini, J. and Gebru, T. (2018) "Gender shades: Intersectional accuracy disparities in commercial gender classification", *Proceedings of Machine Learning Research*, 81, pp. 77–91.

Cabral, L., Pinto, R. and Gonçalves, G. (2025) "AI-powered learning analytics dashboards: a systematic review of applications, techniques, and research gaps", *Discover Education*, 4, 525. doi:10.1007/s44217-025-00964-y.

Crenshaw, K. W. (1989) "Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics", *University of Chicago Legal Forum*, 1989, pp. 139–167.

Cukurova, M. (2025) "The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence", *British Journal of Educational Technology*, 56(2), pp. 469–488. doi:10.1111/bjet.13514.

Drachsler, H. and Greller, W. (2016) “Privacy and analytics: it’s a DELICATE issue. A checklist for trusted learning analytics”, in *Proceedings of the 6th International Conference on Learning Analytics and Knowledge (LAK ’16)*, Edinburgh, UK, pp. 89–98. doi:10.1145/2883851.2883893.

Hooshyar, D., Šír, G., Yang, Y., Kikas, E., Hämäläinen, R., Kärkkäinen, T., Gašević, D. and Azevedo, R. (2025) “Towards responsible AI for education: Hybrid human-AI to confront the elephant in the room”, *Computers and Education: Artificial Intelligence*, 9, 100524. doi:10.1016/j.caeai.2025.100524.

Hossain, R. (2026) *Exploring Intersectional Fairness and Bias Mitigation Techniques in Job Recommendation*. Master’s thesis. University of Koblenz. Available at: https://opus4.kobv.de/opus4-uni-koblenz/frontdoor/index/index/docId/2638 (CC BY-SA 4.0). doi:10.82549/opus4-2638.

Ibrahim, A., Alarood, A. and Alsolami, E. (2025) “An equity aware recommender system for university admissions balancing operational constraints and strategic objectives”, *Scientific Reports*, 15, 39756. doi:10.1038/s41598-025-23116-6.

Mirsch, M., Strube, J. and Leicht-Scholten, C. (2025) “Invisible inequalities - intersectional fairness in educational AI”, *Proceedings of Machine Learning Research*, 294, pp. 403–409.

Sarmurzin, Y., Azhenov, A., Kudysheva, A., Ualikhanova, B., Saparbekova, I. and Besbayev, A. (2026) “Implementation of AI in career counselling for university students: a systematic review”, *Frontiers in Education*, 11, 1787689. doi:10.3389/feduc.2026.1787689.

Shailya, K., Mishra, A.K., Krishnan, G.S. and Ravindran, B. (2025) “Where should I study? Biased language models decide! Evaluating fairness in LMs for academic recommendations”, in *Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*, Mumbai, India, pp. 2291–2317. doi:10.18653/v1/2025.findings-ijcnlp.141.

Simbeck, K. (2024) “They shall be fair, transparent, and robust: auditing learning analytics systems”, *AI and Ethics*, 4, pp. 555–571. doi:10.1007/s43681-023-00292-7.

Slade, S. and Prinsloo, P. (2013) “Learning analytics: ethical issues and dilemmas”, *American Behavioral Scientist*, 57(10), pp. 1510–1529. doi:10.1177/0002764213479366.

Topali, P., Ortega-Arranz, A., Rodríguez-Triana, M.J., Er, E., Khalil, M. and Akçapınar, G. (2025) “Designing human-centered learning analytics and artificial intelligence in education solutions: a systematic literature review”, *Behaviour and Information Technology*, 44(5), pp. 1071–1098. doi:10.1080/0144929X.2024.2345295.

Weissburg, I., Anand, S., Levy, S. and Jeong, H. (2025) “LLMs are biased teachers: Evaluating LLM bias in personalized education”, in *Findings of the Association for Computational Linguistics: NAACL 2025*, Albuquerque, New Mexico, pp. 5665–5713. doi:10.18653/v1/2025.findings-naacl.314.

Zacharis, G. and Papadakis, S. (2025) “Can AI grade like a human? Validity, reliability, and fairness in university coursework assessment”, *Educational Process: International Journal*, 19, e2025591. doi:10.22521/edupij.2025.19.591.