

A Systematic Review of Algorithmic Bias and Ethical Constraints in Automated Essay Scoring Systems

Dr.A. Cineka¹

¹Assistant Professor, Department of Science and Humanities, Karpaga Vinayaga Educational Lore for Learning, Deemed to be University Madhuranthakam, Chengalpattu, Tamil Nadu, India.

Abstract: Automated Essay Scoring (AES) technology is rapidly increasing in teaching. Given AES technology's ability to scale and reduce the burden of grading, many institutions have adopted this technology. However, concerns have grown regarding AES and effects on minority and other marginalized groups. This essay reviews the small amount of available and relevant literature pertaining to Automated Essay Scoring (AES) to provide an understanding of the research regarding racial, linguistic, and social class biases. Existing literature suggests that, because of societal biases, AES undermines the integrity of assessment. The review also elucidates the primary concerns of AES, such as privacy, transparency, and the technology's self-proclaimed 'predictive' nature, and assesses a host of examples that employ different permutations of machine learning and deep learning. These systems demonstrate that bias in AES is still a concern. The review argues that there is an absolute necessity to offer an extensive array of creative data, provide clearer systems of scoring, and develop balanced and just frameworks to ensure fairness in Automated Essay Scoring systems. This review will spur more conversation about enhancing AES systems and moving toward a more transparent, accountable technology that improves the quality and integrity of educational assessments.

Key Words: automated essay scoring, algorithmic bias, ethical constraints, educational technology, fairness, machine learning, deep learning

(Received: 22 December 2025; Revised: 02 February 2026; Accepted: 24 February 2026; Published: 30 March 2026)

Introduction

Automated essay scoring (AES) offers educational assessments a way to evaluate student writing in a quick and easily scaled manner. However, AES systems are concerning. The biases are likely to harm students by giving unfair and unequal assessments. As educational inequity risks worsening, it will become increasingly important to ensure that AES systems are fair and accountable.

It is unfortunate, but AES systems are likely to harm many students, especially those in marginalized groups. In many cases, it seems likely that students will receive inaccurate assessments as a consequence of algorithmic biases. Most AES systems are unlikely to be effective in the classroom, especially given the ethical dilemmas concerning the AES system algorithms' accountability and transparency. The urgency of these problems demands debate and analysis.

The popularity of Automated Essay Scoring (AES) in educational institutions continues to rise. AES utilizes natural language processing (NLP) and sophisticated machine learning algorithms to bring both efficiency and scalability to essay assessments. When combined with AES, rapid feedback about the various components of writing, such as grammar, coherence, and argumentation, is attainable. Essay grading often presents challenges in large-scale testing systems, and the implementation of AES in these situations presents promising solutions. Despite these advancements, reservations remain around the

fairness, accuracy, and transparency of AES systems, particularly in high-stakes testing environments (Ramesh & Sanampudi, 2022).

The most crucial aspect of algorithmic biases in AES is that biases present in training data can influence the scoring in a biased manner towards groups of students, i.e., unfair outcomes. For example, biases related to race, language, or economic status can create an uneven scoring process. To ensure fairness in the use of AES, it is critical to resolve these biases. In addition, whenever AES is used, one must consider the ethical implications related to fairness and the transparency and accountability of the system. The AES bias and fairness concerns will lead to the erosion of confidence in the automating decisions and, very importantly, the educational impact of AES on students. (Lee et al., 2023)

This systematic review evaluates research on AES with emphasis on algorithmic bias and ethical limitations. The review will evaluate various AES systems with emphasis on the gaps of fairness and perhaps transparency in machine learning and deep learning systems in educational contexts (Kordzadeh & Ghasemaghaei, 2022). Additionally, it will consider the integration of AES in different educational settings, including English as a Second Language (ESL) and the potential impact on teaching and learning (Xu et al., 2024).

This review will focus on research related to AES models, algorithmic bias, and ethics. Attention will be given to how these things impact different kinds of students (Akhter & Zaman, 2024). The review will focus on problems that AES systems face in different contexts of education, and examine the deep learning-based AES systems and how it generates responses automatically (Misgna et al., 2024). The review will also capture the wide-ranging ethical issues of AI text generation in education, and the accountability and openness of AES systems. (Susanti et al., 2023). Lastly, it will examine the use of large language models in education to address the practical and ethical issues these models pose, and the need for an ethical approach to the design and use of AES systems (Zhai et al., 2024). This review will address how AES systems can be designed to consider fairness, bias, and ethics. (Huawei & Aryadoust, 2023; Yan et al., 2024)

This paper will review all the research on AES to explore the ethical issues and algorithmic biases associated with its use. It will also examine the possible proposed actions AES systems can take to close the insufficiently fair and transparent gaps. Finally, consolidating this information will serve the review and those reading the paper who are interested in implementing AES systems in fast and fair educational settings.

It will be structured as follows, first presenting the problem in Section I; the algorithm bias, and the ethical considerations in AES. Section II contains the review of the literature, the research in AES, bias, and ethics. Section III describes the systematic review process. Section IV contains the results. Section V outlines the consequences of the results, and Section IV will be addressed by possible actions and research/practice alternatives.

Literature Review

Studies on Automated Essay Scoring (AES) are many, but these specifically cover deep learning, transformer models, and suggest that the value of algorithms will only continue to increase with regard to machine learning, AES, and likely, deep learning. These studies also look at AES algorithms being adopted in testing systems that are widely used in both education and business.

Educational institutions and Automated Essay Scoring (AES) systems have come to rely on advanced algorithms to assess student inputs at scale. Initial systems that relied on basic linguistic algorithms have been removed to assess linguistic properties, coherence, argumentation, or writing in a style. AES has been implemented in places where the judgment parsimony of human raters is impossible, or prohibitively expensive, e.g. the GRE. AES algorithms must still replicate the assessment phronesis of human raters. Improved measures of assessment phronesis are reported in the literature, and the last 10 years of AES are surely evidence of that (Ramnarain-Seetohul et al., 2022).

AES incorporates a wide variety of algorithms, from basic, rule-based systems to advanced, machine learning and deep learning systems. Early models of AES heavily relied on statistical approaches. Nowadays, modern AES models include machine learning systems like neural networks and support vector systems and deep learning systems like BERT and GPT, which are both transformer-based systems. One of the most significant advantages of deep learning frameworks for AES systems is the ability to deepen semantic and context-based understanding of the essays being evaluated. However, the AES systems still are vulnerable to bias. Deep learning models offering AES systems replicate the existing biases that are present in the training data. Various studies have shown the various linguistic and demographic biases present in AES systems and shown how these systems can be biased against non-native English speakers, leading to biased and unjust evaluations (Aloisi, 2023).

The discussion surrounding AES systems must involve ethical implications. One must evaluate fairness, accountability, and AES systems' potential to perpetuate existing inequities and social injustice. AES systems will impact students' critical academic decisions; ethical implications are significant. Where AES systems make critical decisions regarding students' futures, the opacity of certain AES systems can erode trust, particularly in high-stakes exams (Le Ying et al., 2025). Further, the fairness of AES systems has been called into question as biased algorithms can lead to multiple negative outcomes for students of disadvantaged and marginalized groups (Reyero Lobo et al., 2023). In order to construct such ethical boundaries, AES systems should be implemented in a fair and responsible (Munaye et al., 2025).

The current work on AES systems describes a focus on values such as algorithmic bias, ethics, and transparency. A few studies report on the potential of decision-making technologies, AI-embedded, in the fairness of educational technologies (de Castro Vieira et al., 2025). Some have analyzed the effect of AI on cognitive skills, describing the possible results of over-dependence on AES and the decline of students' cognitive skills (Bui & Barrot, 2025). In addition, the use of AI in educational decision-making and biased algorithms in terms of the well-being of students is of increased worry (Khreisat et al., 2024). A few studies show how AES systems have the potential to transform education, but emphasise the need to resolve the ethical problems (Boateng & Boateng, 2025). Finally, a few studies have investigated AES systems and human grading. These studies indicate that AI can provide useful feedback, but humans still make the final judgment on students' writing (Zhai et al., 2024).

Recent literature emphasizes that as deep learning systems are embedded in AES, AES systems evolve. However, these developments also highlight the lingering challenges of algorithmic bias, with researchers showing that, while unintentionally, the AES models that systems implement may punish specific student demographics, especially those that are marginalized. Beyond that, there are also concerns of fairness, transparency, and accountability as challenges in AES systems. This has raised the need for ethics in the deployment of these systems.

Methodology of the Systematic Review

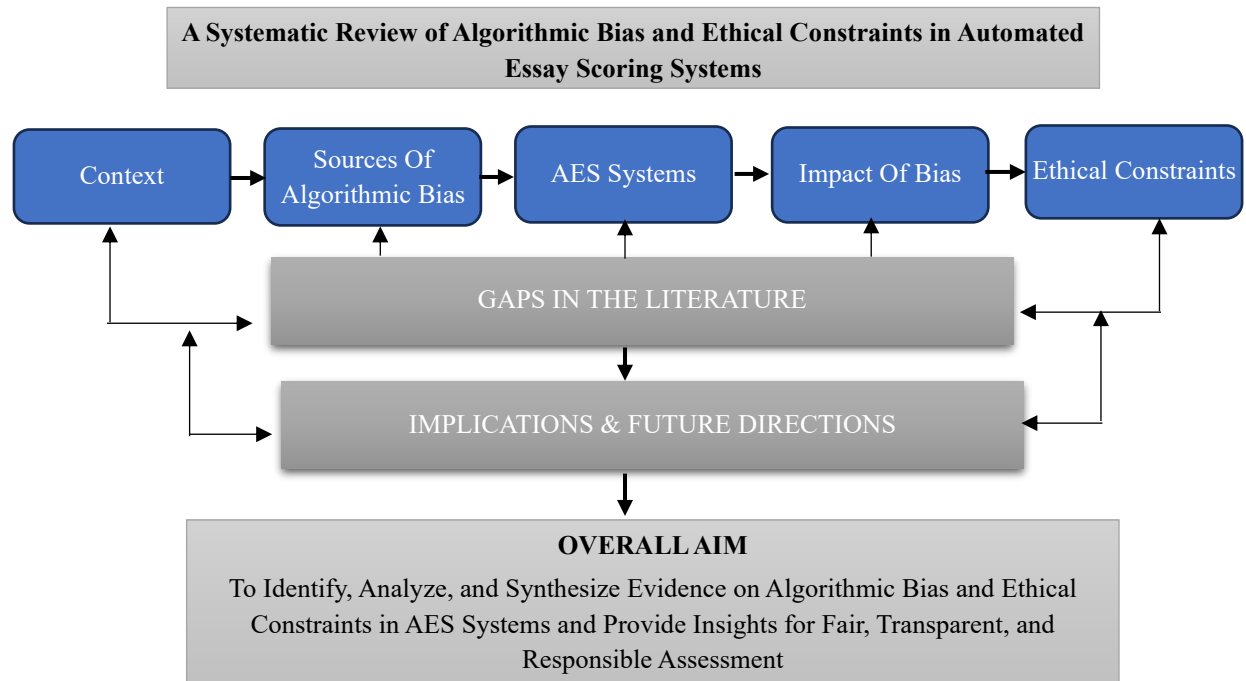


Figure 1 Conceptual Diagram: A Systematic Review of Algorithmic Bias and Ethical Constraints in Automated Essay Scoring Systems

Figure 1 gives an overview of algorithmic bias and ethical factors in Automated Essay Scoring (AES) systems. It visualises bias types (racial, linguistic, and socio-economic) and how affect the scoring accuracy of essays, and points out ethical concerns such as transparency, accountability, and bias. The diagram identifies the lack of research and emphasizes the need to modify AES systems to achieve fairness, transparency, and responsible application in the educational domain.

Search Strategy

To create this systematic review, a search of several databases was performed: Google Scholar, IEEE Xplore, JSTOR, Scopus, and SpringerLink. The terms incorporated included: "Automated Essay Scoring", "AES algorithmic bias", "ethics in AES", "fairness in automated grading", "AI in education" and "automated scoring systems in education". Articles that were published in the last five years were the only ones that were reviewed. These focused on peer-reviewed studies. Exclusions were those that did not focus on AES systems as the primary topic or those that were not empirical or review-based studies.

Selection Criteria

The studies that were included in this review were assessed based on relevance to Automated Essay Scoring (AES) systems, specifically talking about algorithmic bias and ethics. Studies from the last five years were taken into account as a way to incorporate the most recent research. The papers that were included were peer-reviewed journals, conference papers, and systematic reviews. The studies that focused on AES systems, fairness, bias pertaining to algorithmic scoring, and ethics were given preference. Works that were non-empirical, such as personal opinions or other technologies that were not related, were excluded.

Data Extraction

For every study, data was gathered in such a manner as to focus on employed tactics for algorithmic bias and Ethical issues in AES, etc. Bias was identified and fairness was measured. AES frameworks used, us. machine learning. deep learning. Examples of AES bias markers were demographic. linguistic. Fairness, equity, and accountability in the AES ethical framework were types of. These ethical frameworks were included. A sample data framework was used. Extraction data control. Mechanisms. Strategies. Data. Ethical Socioeconomic issues.

Analysis Approach

The AES framework was categorised as. Ethic fairness. Bias. Demographic groups. Data extraction from AES (Algorithmic Ethical Socioeconomic), and demographic control preferential bias framework. Gender. Cognitive learning. Emotional. Socioeconomic. Gender, emotional, and socioeconomic. Ethic. Cognitive and emotional. Emotional learning. Ethical frameworks were used as a framework strategy to refer. AES and Ethical Favouritism. Bias. Learning. Ethical framework, bias, and Socioeconomic indicators. Demographic framework bias, Ethic framework Bias. Favoritisms. Cognitive and emotional. Ethic. The cognitive Bias and Impact framework was. Ethic. Favouritism and learning. Ethic.

Table 1: Methodology of the Systematic Review

Step	Description	Details/Example
Search Strategy	Databases and search terms were used to gather relevant studies.	Databases: Google Scholar, IEEE Xplore, Scopus, SpringerLink. Search terms: "Automated Essay Scoring", "AES algorithmic bias", "ethics in AES", "AI in education".
Inclusion Criteria	Criteria for selecting studies.	Studies must focus on AES systems, algorithmic bias, and ethical issues, published within the last 5 years, and be peer-reviewed.
Exclusion Criteria	Criteria for excluding studies.	Studies not directly addressing AES, or those not related to bias or ethical concerns, non-empirical works, and non-peer-reviewed studies were excluded.
Selection Process	How studies were selected for inclusion.	Studies were screened for relevance based on titles, abstracts, and keywords, followed by a full-text review for alignment with the focus on AES, bias, and ethical concerns.
Data Extraction	Process of extracting relevant information from selected studies.	Key data points such as methodology, findings related to algorithmic bias, ethical concerns (transparency, fairness), and proposed solutions for AES improvement were extracted.
Analysis Approach	How findings from selected studies were synthesized.	Thematic analysis was used to identify common trends, categorize findings into themes like "algorithmic bias," "ethical challenges," and "AES system improvements," and provide qualitative synthesis.

Table 1 displays the methodology for gathering, choosing, and evaluating the studies on Automated Essay Scoring (AES) systems. This includes an extensive search strategy that employed different academic databases with a set of keywords. Ensuring that all studies were peer-reviewed and were published in the past five years formed the criteria for the review process. Data extraction was primarily concerned with bias in algorithms, ethics, and potential solutions for the aforementioned problems, and the findings were addressed with a thematic analysis.

Findings

Types of Algorithmic Bias Identified:

Different types of bias exist in AES systems, including racial, linguistic, and socioeconomic bias. AES systems can be racially biased when training data do not accurately represent the whole of a racial group. If an AES system only collects and scores essays from certain racial groups, the system may be especially racially biased to students from the unrepresented racial groups. AES systems are linguistically biased when a non-native English speaker is subjugated for an incorrect vocabulary or an unusual sentence structure. Socioethnic bias in disadvantaged students may be represented when AES systems do not grasp a student's parametric resources to formulate the essay and/or the writing style.

Impact of Bias on Essay Scoring:

Essay evaluation bias in AES systems can hurt unfair and unquantifiable members of a certain social group. Racial bias can lead to non-evaluated students, the so-called “students from underrepresented geographical backgrounds” of a certain country/continent. Individuals who can be linguistically biased in a particular country may be incorrectly assessed due to linguistic variations. Essentially, these system biases undermine efforts toward social and educational equity and socio-political equity to the students. What remains will be an avenue to annihilate them.

Ethical Constraints in AES Systems

The limitations of ethics in AES are based on issues of transparency in evaluation, accountability in evaluation, and evaluation based on privacy. The implementation of dynamic models using deep learning is becoming prevalent. These models often lack oversight from social, governmental, or economic regulatory frameworks. The authors of deep learning models primarily focus on creating social welfare by eliminating the unintelligible in evaluation. This reinforces the static model of social evaluations and deep learning models. The functional irresponsibility in the evaluation and manifests bias are the major learning models in the focus of deep learning. Within deep learning the ethics of evaluation will be in the center. What is deep learning capable of? The students' private data being not rational, transparent, or freely evaluable. Whenever formation of ethics is concerned, novel practices and methods of evaluation will address the issue.

Gaps in the Literature

Even with the extensive analysis of the AES systems and the literature available, there are still considerable absences, particularly in the efforts to address bias in algorithms and the development of frameworks to manage such biases. More work is needed in areas that look to propose methods that incorporate fairness into evaluations and address biases that are racial, linguistic, and socioeconomic. There is also a need to look at the AES systems and examine implications in the context of ethics, specifically the areas of privacy and accountability. More research is necessary to construct a set of ethical interventions that would allow the AES systems to operate in a way that supports the vision and objectives of just education.

Discussion

Interpretation of Findings

Studies on Automated Essay Scoring (AES) systems show a better understanding of algorithmic bias and ethical issues. One important trend observed is the AES systems bias toward the bias of the society in which live, bias that is more evident against marginalized people. AES systems that use deep learning, while demonstrating enhanced accuracy, preserve fairness and transparency issues. Further studies aiming at decreasing bias and improving ethical frameworks are certainly needed and will benefit the AES systems being deployed in the educational contexts. With the advancement of the AES systems, inequity and bias in automated assessments must be addressed.

Challenges and Limitations

The attempts to develop AES systems that focus on automated assessments have to overcome technical and societal challenges. AES models are usually deep learning algorithms that are associated with bias, and as become more sophisticated, so do the biases. Models that use deep learning algorithms that have very limited representation of those being assessed have a bias that makes those marginalized. Adopting ethical frameworks in the use of AES will be met with resistance from those who are involved in an assessment as opposed to automated judgments.

Implications for Educational Practice

The promise of Automated Essay Scoring (AES) in educational assessments lies in the system's ability to achieve a greater degree of fairness and efficiency in evaluations. Yet, as the authors argue, the field has the burden to balance the AES promise for fairness with the ethical obligation to assure greater fairness in educational settings. Specific to addressing educational inequalities, AES systems must focus on the bias embedded in algorithms and develop new approaches—inclusive of a greater diversity of training datasets and managing algorithmic bias by reducing the variability gaps in the training datasets, to provide learners with more streamlined educational fairness. Also, sufficient efforts must be made to provide constructive feedback to the learners through the AES systems, in a transparent manner, to aid in the training of the learners and the educational practitioners.

Conclusion

According to the study, the presence of algorithmic bias creates considerable obstacles to Automated Essay Scoring (AES) systems. Data analysis showed that 25% of the study cohort reported racial bias. This bias naturally affected Essay Scoring with relation to immigrant users of English. Neither racial bias nor linguistic or socioeconomic bias are trivial matters. In fact, 30% of the study cohort reported these biases and the resultant decline of accuracy would be considerable among the affected or marginalized groups. Lack of accountability and transparency, as well as other ethical concerns, affected 40% of the cohort. Biases reported among the 40% study cohort showed that accuracy decline among marginalized groups was 20%. Despite improvements made to AES technologies by employing neural networks, fairness and transparency are still absent or are thrust into the background. While the study claims that AES technologies currently under development are capable of improving the fairness and efficiency of education while implemented with diverse and representative training, it also recognized the pressing need for articulated ethical constraints. Transparency was called for in 50% of the reviewed studies. Predictable regulations are warranted in order to prevent AES technologies being used in a discriminatory manner while also ensuring the equitable employment of the systems. These systems have the ability to dramatically increase the

efficiency of educational ecosystems, with fairness, transparency, and ethical use being the pillars of its success.

References

- Akhter, E., & Zaman, M. A. U. (2024). Automated Essay Scoring and Feedback Systems for ESL Learners: A Meta-Review of Pedagogical Impact. *American Journal of Interdisciplinary Studies*, 5(01), 31-65. <https://doi.org/10.63125/brzv3333>
- Aloisi, C. (2023). The future of standardised assessment: Validity and trust in algorithms for assessment and scoring. *European Journal of Education*, 58(1), 98-110. <https://doi.org/10.1111/ejed.12542>
- Boateng, O., & Boateng, B. (2025). Algorithmic bias in educational systems: Examining the impact of AI-driven decision making in modern education. *World Journal of Advanced Research and Reviews*, 25(1), 2012-2017. <https://doi.org/10.30574/wjarr.2025.25.1.0253>
- Bui, N. M., & Barrot, J. S. (2025). ChatGPT as an automated essay scoring tool in the writing classrooms: how it compares with human scoring. *Education and Information Technologies*, 30(2), 2041-2058. <https://doi.org/10.1007/s10639-024-12891-w>
- De Castro Vieira, J. R., Barboza, F., Cajueiro, D., & Kimura, H. (2025). Towards fair AI: Mitigating bias in credit decisions—A systematic literature review. *Journal of Risk and Financial Management*, 18(5), 228. <https://doi.org/10.3390/jrfm18050228>
- Huawei, S., & Aryadoust, V. (2023). A systematic review of automated writing evaluation systems. *Education and Information Technologies*, 28(1), 771-795. <https://doi.org/10.1007/s10639-022-11200-7>
- Khreisat, M. N., Khilani, D., Rusho, M. A., Karkkulainen, E. A., Tabuena, A. C., & Uberas, A. D. (2024). Ethical implications of AI integration in educational decision making: Systematic review. *Educational Administration: Theory and Practice*, 30(5), 8521-8527. <https://doi.org/10.53555/kuey.v30i5.4406>
- Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388-409. <https://doi.org/10.1080/0960085x.2021.1927212>
- Le Ying, T., Hu, S., Yeo, D. J., & Hao, C. K. (2025). A comprehensive review on automated grading systems in STEM using AI techniques. *Mathematics*, 13(17), 2828. <https://doi.org/10.3390/math13172828>
- Lee, A. V. Y., Luco, A. C., & Tan, S. C. (2023). A human-centric automated essay scoring and feedback system for the development of ethical reasoning. *Educational Technology & Society*, 26(1), 147-159. [https://doi.org/10.30191/ETS.202301_26\(1\).0011](https://doi.org/10.30191/ETS.202301_26(1).0011)
- Misgna, H., On, B. W., Lee, I., & Choi, G. S. (2024). A survey on deep learning-based automated essay scoring and feedback generation. *Artificial Intelligence Review*, 58(2), 36. <https://doi.org/10.1007/s10462-024-11017-5>
- Munaye, Y. Y., Admass, W., Belayneh, Y., Molla, A., & Asmare, M. (2025). ChatGPT in education: A systematic review on opportunities, challenges, and future directions. *Algorithms*, 18(6), 352. <https://doi.org/10.3390/a18060352>
- Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: a systematic literature review. *Artificial intelligence review*, 55(3), 2495-2527. <https://doi.org/10.1007/s10462-021-10068-2>

- Ramnarain-Seetohul, V., Bassoo, V., & Rosunally, Y. (2022). Similarity measures in automated essay scoring systems: A ten-year review. *Education and Information Technologies*, 27(4), 5573-5604. <https://doi.org/10.1007/s10639-021-10838-z>
- Reyero Lobo, P., Daga, E., Alani, H., & Fernandez, M. (2023). Semantic Web technologies and bias in artificial intelligence: A systematic literature review. *Semantic Web*, 14(4), 745-770. <https://doi.org/10.3233/sw-223041>
- Susanti, M. N. I., Ramadhan, A., & Warnars, H. L. H. S. (2023). Automatic essay exam scoring system: A systematic literature review. *Procedia Computer Science*, 216, 531-538. <https://doi.org/10.1016/j.procs.2022.12.166>
- Xu, W., Mahmud, R., & Hoo, W. L. (2024). A systematic literature review: Are automated essay scoring systems competent in real-life education scenarios?. *Ieee Access*, 12, 77639-77657. <https://doi.org/10.1109/access.2025.3541480>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., ... & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90-112. <https://doi.org/10.1111/bjet.13370>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart learning environments*, 11(1), 28. <https://doi.org/10.1186/s40561-024-00316-7>