Online Learning in Educational Research

Volume 5, Issue 1, 189-216.
e_ISSN: 2798-2351
https://www.journal.foundae.com/index.php/oler/index
DOI: 10.58524/oler.v5i1.695

Unveiling the Dual Nature of AI in Grading: A Systematic Review of Benefits and Mitigation Strategies for Algorithmic Bias

Eko Risdianto*
Universitas Bengkulu
INDONESIA

Rince Aida Rostika
Universitas Bengkulu
INDONESIA

Ahmad Zabidi Abdul Razak
Universiti Malaya
INDONESIA

Hera Anis Kartika
Universitas Bengkulu
INDONESIA

Jeni Fitria
Universitas Bengkulu
INDONESIA

Article Info

Article history:

Received: January 17, 2025
Revised: April 20, 2025
Accepted: May 19, 2025

Keywords:

AI in Education;
Artificial Intelligence;
Automated Assessment;
Bibliometric;
Systematic Literature Review.

Abstract

The use of Artificial Intelligence (AI) in educational evaluation optimizes learning outcomes. This research seeks to address the advantages and difficulties of implementing AI within academic evaluation frameworks with particular emphasis on the algorithmic bias problem and its implications for fairness in education. The absence of a thorough grasp of algorithmic bias, particularly how it can be utilized as a weapon against equitable education, reveals an important gap. We conduct a Systematic Literature Review (SLR) and bibliometric analysis on 121 articles sourced from Scopus published between 2021 to 2025 to trace the trends and examine the impacts and biases of AI on grading systems. The data demonstrates a significant increase in publications beginning 2018, concentrating on topics such as educational applications of AI, automated grading systems, and machine learning. The findings further indicate that though AI improves efficiency and consistency of the evaluations, it heightens the chances of biased outcomes because of non-diverse training data, prejudiced developers, and socio-cultural frameworks that could worsen the situation for already marginalized learners. In summary, this study highlights the critical gaps in bias mitigation strategies arising from the lack of ethical design frameworks, antecedent-free algorithms, and educator prep courses aimed at combating bias. These outcomes serve as benchmarks for the creation of more reliable and comprehensive AI systems for assessments and shift subsequent investigations to focus validation on different cultures and the incorporation of just AI design paradigms.

To cite this article: Risdianto, E., Rostika, R. A., Razak, A. Z., Kartika, H. A., & Fitria, J. (2025). Unveiling the dual nature of ai in grading: A systematic review of benefits and mitigation strategies for algorithmic bias. *Online Learning in Educational Research, 5*(1), 189-216. https://doi.org/10.58524/oler.v5i1.695

INTRODUCTION

The adoption of pioneering technologies, particularly artificial intelligence (AI), has greatly changed numerous fields, and education has received perhaps the deepest impact. It fosters new and critical methods of teaching and learning while simultaneously enhancing the quality of education at every level (Branch, 2009; Kamalov et al., 2023 Sasilatha & Suprianto, 2025). The transformation helps us to more advanced methods of teaching and learning, as well as a more tailored form of education, resulting in higher quality education (Kasim et al., 2025; Vieriu & Petrea, 2025). The implementation of AI in education technology has also achieved considerable improvements in engagement, self-paced learning, and teacher aids (Luckin et al., 2022). Managing

* **Corresponding author**
Eko Risdianto Universitas Bengkulu, Indonesia. eko_risdianto@unib.ac.id

the learner's experiences is one of the advantages AI technologies provide. In the current environment, AI can serve as an assistant for both learners and educators (Armoogum & Zakaria, 2024; Herawati et al., 2024). AI is capable of analyzing individual performance metrics and adequately customizing learning materials to suit the various levels within a class. The implementation of AI technology into education has the potential to customize educational pathways, thereby enhancing learning achievement (Aghaziarati et al., 2023). Integrating technology into classrooms enhances the efficiency of conventional methods. The application of artificial intelligence allows each student's learning journey to be tailored according to their particular strengths and weaknesses (Khosravi et al., 2023). The more aware a teacher is about AI, the more likely they are to use this technology in the classroom (Aghaziarati et al., 2023).

AI helps simplify the assessment process, which typically requires much effort and time. This technology enables schools and universities to automatically grade large numbers of student assignments, significantly increasing the efficiency of assessment and allowing students to receive timely feedback on their performance (Crompton & Burke, 2023). While AI can help with assessment, its use raises ethical and emotional concerns for students and teachers. Many studies show that students feel uncomfortable if AI replaces traditional assessment methods, as they worry that this technology cannot assess with the deep understanding that teachers do(Khater et al., 2023). Moreover, too much reliance on AI for assessment can jeopardize academic integrity. This approach poses a threat to a more sophisticated understanding of students' responses, resulting in grade inflation due to the misinterpretation of complex student answers (Coghlan et al., 2021).

One of the key advantages of AI in assessment is its ability to improve efficiency and consistency, which is very important in today's education. For example, AI technology can grade essays by analyzing grammar, structure, and content, thus reducing the administrative burden on educators (Owan et al., 2023). However, AI uses the same criteria in grading all assignments, so it can help make student assessments fairer and objective. Relying too much on AI in assessment can make the learning experience feel less human. Students may feel that their assignments are being graded by machines, not by someone who truly understands their learning process (Opesemowo & Adekomaya, 2024). It can lead to a lack of motivation and student engagement, as direct interaction with teachers is vital in creating a supportive learning environment. AI that provides standardized feedback may not be able to adapt to each student's unique needs, even though personalization is essential in the learning process (Damaševičius, 2024). This situation is further complicated by the possibility of AI encouraging dishonest academic actions. If they rely too much on automated assessments, students may be tempted to look for ways to manipulate the system to get high grades, so they consider shortcuts acceptable (Crompton & Burke, 2022). These concerns emphasize the need for educators to maintain a balanced approach by ensuring that AI complements but does not replace their pedagogical expertise.

The integration of AI technology into education systems is expanding, particularly in evaluating students. However, this raises issues of injustice and discrimination, which could disadvantage students from different backgrounds. AI systems tend to incorporate preexisting bias within their datasets (Zong et al., 2023). Furthermore, the application of AI for adaptive instruction has the potential to worsen inequality disproportionally. The objective is to modify and optimize the educational content for every learner. In reality, it only aids privileged learners, thereby increasing the divide with the underprivileged (J. Lee et al., 2024). One might attribute this phenomenon to education, where students endowed with AI will be better equipped to exploit it.

Consequently, there is a disparity in evaluating students' skills and talents, which may foster biased views among diverse social strata (Anuyahong et al., 2023). This scenario is further exacerbated by bias stemming from inequality concerning the evaluation of students. Gender, ethnicity, and even nationality can influence how students are appraised, which leads to discrimination within the evaluation framework (F. Kim et al., 2024). To tackle these issues, educators and other stakeholders involved in implementing AI should be provided with clear ethical boundaries and comprehensive training related to AI use in Education. Educators and school leaders need to comprehend the biases that can arise from algorithms so that efforts may be taken to mitigate algorithmic social injustice (Gándara et al., 2024). Furthermore, bias detection and mitigation in educational algorithms should be systematically designed in order to increase fairness within the evaluation processes in education (Baker & Fairclough, 2021). In conclusion, although AI has the potential to enhance accuracy and efficiency in educational assessments, ingrained biases pose a significant challenge to the attainment of fairness and equality in academic institutions. Addressing this challenge requires collaborative understanding, decisive action, and dedicated efforts towards achieving equity in the use of AI in education.

Numerous problems can stem from the use of AI in education that will likely change fairness among students when assessing teachers. A lack of representation among all students in the training data is one of the foremost issues (Pham et al., 2025). Many AI systems are trained with data that only covers certain groups, leaving them far from reflecting the diversity present in the educational landscape. AI can either help or hinder the evaluation of a group of students. Additionally, the algorithms used in the evaluation of AI systems could generate bias. Bias occurs when an algorithm reflects the beliefs of its programmers, often made without conscious intent. For example, if an algorithm is created based on specific "ideal" benchmarks set by a programmer, there could be students who are capable of performing better than those who are deemed ideal but do not meet the programmed standards (Tri & Nataliani, 2021). As for social and cultural elements, peer views and general sociological attitudes towards education and assessment can shape the algorithm's evaluation metrics and student scoring frameworks (Ramadan, 2024). Overall, the more structured the approach to mitigating bias in assessments, the more effective it will be. Suggested approaches outline wider use of data, refinement of the algorithm, and deeper appreciation of possible biases within the system by developers and users.

Reflecting on technological innovations and ethical dilemmas in education, targeted research on the impact AI technology has on assessing a student's academic performance conducted between 2010-2024 is nothing short of fascinating. Analyzing the advantages and biases of AI in academic assessment between the years 2010 and 2024 shows an ongoing process characterized by rapid technological development and persistent ethical concerns. Past research has noted the extensive capabilities of AI to assist with personalized education, as well as the algorithmic biases that threaten equality within education. However, there are still gaps in addressing these biases concerning culture and current AI technologies. As a response, the most recent studies highlight the need to address the issues of fairness and inclusivity within AI-driven assessment tools in relation to automated algorithms that adapt to students' learning routines in order to provide personalized feedback and foster equitable education in a shifting digital landscape. The studies reveal that AI can be an effective tool for personalizing learning and assessment. These insights allow educators to devise tailored instructional strategies for every learner (Anuyahong et al., 2023). The studies reveal that AI can be an effective tool for personalizing learning and assessment. Educators, guided by AI's analytical capabilities, can more efficiently identify each student's unique strengths and weaknesses (Ifenthaler et al., 2024). Alongside these advantages, the impact of AI on education also raises the concern of bias.

One of the most widespread is bias in algorithms, which usually arises from training data that does not capture all groups of learners. Consequently, AI systems are capable of making unjust assessments, particularly when fairness does not factor into the algorithm's architecture (Zong et al., 2023). Evidence suggests that a lack of diversity within training datasets can heighten bias within AI-based scoring systems (Emilio, 2024). Studies done in the last ten years strongly advocate for the need for policies and practices that uphold and promote transparency and accountability in the application of AI algorithms. Bias can be mitigated with the application of "fair AI," which centers ethical benchmarks in the design and application of AI frameworks. It involves deep impact analysis concerning the use of AI on education so that outcomes reflect fairness and equity (Cavique, 2024). It seeks to understand how educational outcomes are impacted by the technology, ensuring that the technology acts to enhance fairness and equity in student evaluations. As a whole, studies conducted over the past decade provide evidence that although there is great potential for AI to transform grading systems, issues of bias and fairness in AI still pose a challenge. It is hoped that further research can develop a more ethical and equitable way of using AI in the world of education, so that all students can enjoy the benefits of this technology without discrimination.

METHOD

This research follows the Systematic Literature Review (SLR) approach. For the study "AI and Grading Mechanisms: A Critical Review of Benefits and Biases", data was gathered using eight key search terms. The entire dataset was sourced from the Scopus database, and it was filtered to include publications from the years 2021 to 2025. This range was selected because finding recent articles pertinent to the subject matter presented a challenge.

The search terms are grouped into various sets. The first and second keywords include "Artificial Intelligence" OR "AI" OR "Machine Learning" and "Automated Grading" OR "Automated Assessment" OR "AI Grading" OR "Automated Scoring". The fourth to sixth keywords are aimed at finding articles that discuss the benefits as well as biases in the application of AI for assessment, using keywords such as "Artificial Intelligence" OR "AI" OR "Machine Learning", "Automated Grading" OR "AI Grading" OR "Automated Assessment", and "Bias" OR "Fairness" OR "Equity" OR "Ethical Issues". Meanwhile, the seventh and eighth keywords were used to explore aspects of the validity and reliability of AI-based assessment systems, using keywords such as "Artificial Intelligence" OR "AI" OR "Machine Learning", "Educational Assessment" OR "Assessment System" OR "Automated Grading", and "Validity" OR "Reliability" OR "Transparency". After using this keyword, we will get the number of articles we want. Then, all the article data that has been downloaded in CSV format is combined into one file. Furthermore, the merged CSV file is processed using the OpenRefine application. This application cleans data from duplicates and biased articles. After cleaning, the number of valid and ready articles is analyzed.

The next stage is interpreting bibliometric data using the Biblioshiny application. Through this application, various important information is obtained such as keyword metadata, key information on data sheets, graphs and growth tables of articles per year, graphs and tables of citations per year, Three-Field Plot diagrams, list of the most relevant sources, local impact of sources, core sources based on Bradford's Law, cumulative frequency of occurrences, production of authors over time, distribution of correspondent authors by country, as well as graphs and tables related to the scientific contributions of countries in the field (Ullah et al., 2023).

Next, the data is visualized using the VOSviewer application, which displays relationships between articles, authors, and topics in Overlay Visualization and Network Visualization (Yakir et al., 2023). This visualization is very helpful in understanding the patterns of linkages in the literature analyzed. Figure 1 is a diagram of the flow chart of this SLR research.

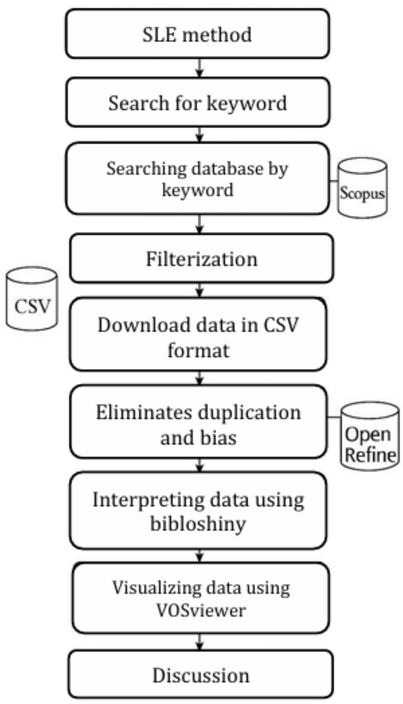


Figure 1 Research Method Flow Diagram

PRISMA Method

The PRISMA method is used to carry out the following data screening process, which is the step used in the method shown in Figure 2.

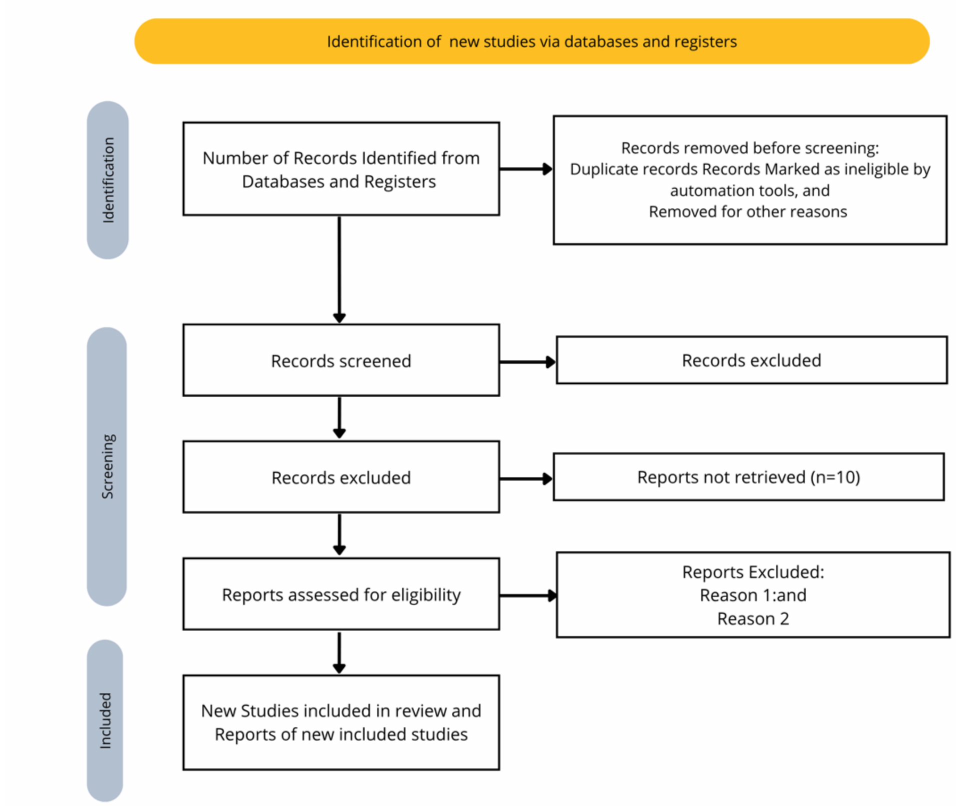


Figure 2. The PRISMA Method

Figure 2, the process of selecting articles for this systematic review is carried out according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and is described in detail through the flow diagram in the PRISMA Figure above. This process consists of four main stages: identification, screening, eligibility, and inclusion.

RESULTS AND DISCUSSION

Bibliometric Studies

The search strategy used Scopus data sources related to this research theme in the bibliometric analysis. From several keywords selected, three keywords were chosen: keywords for a broad search about AI in grading, focusing on the benefits and biases of AI in grading, and keywords to highlight the validity and reliability of AI in academic assessment**.** For a broad search on AI in grading, see Table 1.

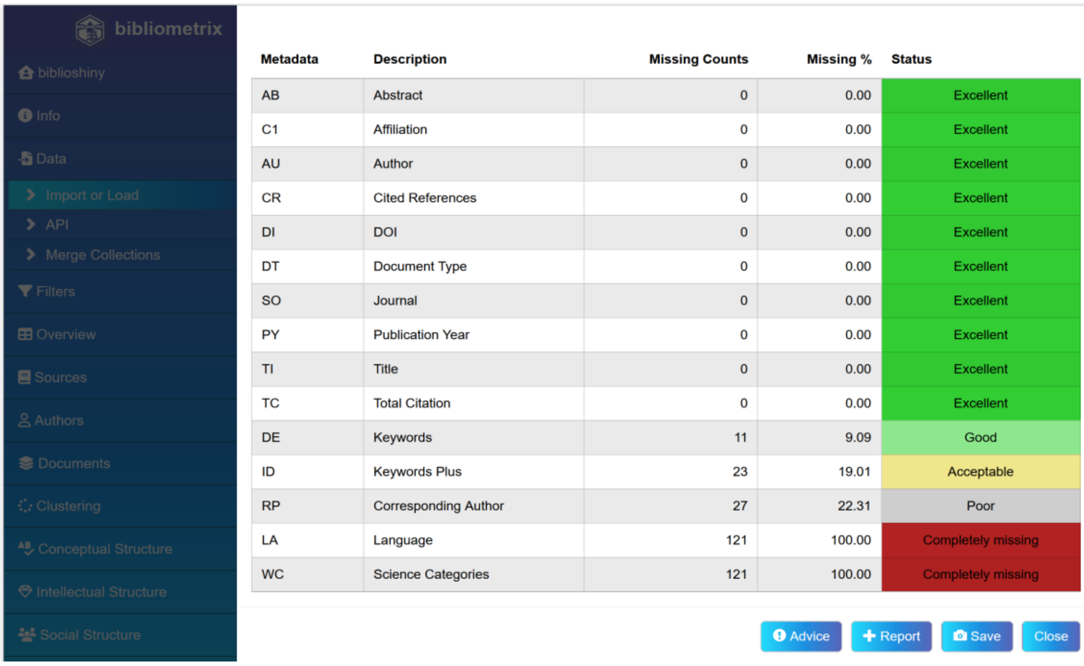
Table 1. Article Data Recapitulation Based on AI Keyword Combination and Automatic Assessment

Keyword Combination	Initial Data Count	Year Range	Filter Criteria	Data Count After Filtering
("Artificial Intelligence" OR "AI" OR "Machine Learning") AND ("Automated Grading" OR "Automated Assessment" OR "AI Grading" OR "Automated Scoring")	349	1997-2025	Limited Article, Open Access, English Language, Article Type, Review, Conference Proceedings	110
("Artificial Intelligence" OR "AI" OR "Machine Learning") AND ("Automated Grading" OR "AI Grading" OR "Automated	2	2015-2025	Limited Article, Open Access, English Language, Article Type,	2

Keyword Combination	Initial Data Count	Year Range	Filter Criteria	Data Count After Filtering
Assessment") AND ("Bias" OR "Fairness" OR "Equity" OR "Ethical Issues")			Review, Conference Proceedings	
("Artificial Intelligence" OR "AI" OR "Machine Learning") AND ("Educational Assessment" OR "Assessment System" OR "Automated Grading") AND ("Validity" OR "Reliability" OR "Transparency")	15	1997-2025	Limited Article, Open Access, English Language, Article Type, Review, Conference Proceedings	14
Total Articles After Merging and Deduplication				126

Reduce duplicate data

Using data from the three files of the search results of the keywords above that are exported, it is compressed into a zip file and then entered into Open Refine. This application filters data, especially to see whether or not the data obtained from search results is there (Pranckutė, 2021). In addition, this application can also be used to eliminate keyword bias in the data received. Initially, the data was obtained from as many as 128 articles, but when entered into Open Refine, it was reduced to 121 articles, as seen in Figure 3.



Metadata	Description	Missing Counts	Missing %	Status
AB	Abstract	0	0.00	Excellent
C1	Affiliation	0	0.00	Excellent
AU	Author	0	0.00	Excellent
CR	Cited References	0	0.00	Excellent
DI	DOI	0	0.00	Excellent
DT	Document Type	0	0.00	Excellent
SO	Journal	0	0.00	Excellent
PY	Publication Year	0	0.00	Excellent
TI	Title	0	0.00	Excellent
TC	Total Citation	0	0.00	Excellent
DE	Keywords	11	9.09	Good
ID	Keywords Plus	23	19.01	Acceptable
RP	Corresponding Author	27	22.31	Poor
LA	Language	121	100.00	Completely missing
WC	Science Categories	121	100.00	Completely missing

Figure 3. The Meta Data on DE/Keyword is Good

Furthermore, the process of interpreting bibliometric data can be carried out using the Bibliosiny application. Table 2 provides the main information on the data as a whole.

Table 2. Main Information on the Data Sheet

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2016:2025
Sources (Journals, Books, etc)	91
Documents	121
Annual Growth Rate %	22.03
Document Average Age	2.83
Average citations per doc	11.83
References	5236
DOCUMENT CONTENTS	

Description	Results
Keywords Plus (ID)	1256
Author's Keywords (DE)	419
AUTHORS	
Authors	635
Authors of single-authored docs	6
AUTHORS COLLABORATION	
Single-authored docs	6
Co-Authors per Doc	5.58
International co-authorships %	22.31
DOCUMENT TYPES	
article	82
conference paper	32
review	7

Table 2 summarizes key information from the research data from 2016 to 2025. In this period, it was recorded that 121 documents were used, coming from 91 different sources such as journals, books, and others. The annual growth rate of publications reaches 22.03%, a relatively rapid increase from year to year. The average age of documents is around 2.8 years, which indicates that most documents are still relatively new. Interestingly, each document gets an average of about 12 citations, indicating that the results of this research are quite influential in their field. This is categorized as quite high. The total references used in the document reached 5,236, indicating that each document referred to multiple sources and had a solid foundation.

Regarding content, it found about 1,256 "Plus" keywords (usually added by the system to expand the topic), and 419 keywords that came directly from the author. This shows that the scope of the research topic is quite broad, but there is still a specific focus according to the author's interests. Regarding collaborations, 635 authors were involved in the overall publication. Only six documents were written by one person, so cooperation is common in this study. On average, there are 5 to 6 authors in one document. In addition, about 22% of the publications involved cooperation between countries, which indicates international involvement in developing this research.

When viewed from the type of documents, most of them are in the form of articles (82 documents), followed by conference papers (32), and the rest in the form of literature reviews (7). This shows that research results are more published in scientific articles and conference presentations than in literature reviews. Overall, these data illustrate that over the past decade, publications in this field have been evolving, are largely collaborative, have a wide range of topic coverage, and have had a considerable impact on the scientific community. The graph in Figure 4 provides a visual overview of how these publication trends have evolved from year to year.

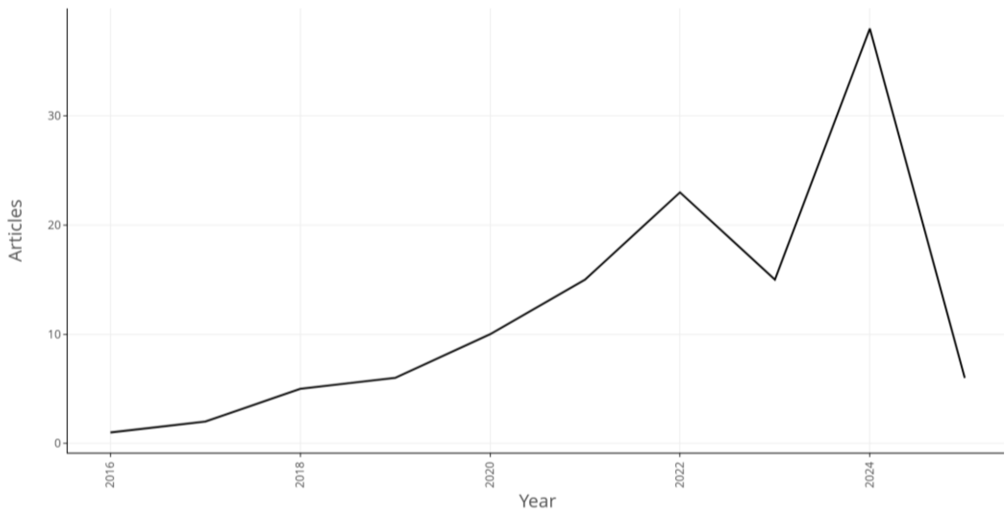


Figure 4. Year-Over-Year Growth Chart

It can be shown by the distribution of the data in Table 3.

Table 3. Yearly Growth

Year	Articles
2016	1
2017	2
2018	5
2019	6
2020	10
2021	15
2022	23
2023	15
2024	38
2025	6

Based on Figure 4 and Table 3, we can see an increasing trend in scientific publications yearly. At the beginning of the period, precisely in 2016, the number of articles published was still very small, only one or two. However, from 2017 to 2021, there was steady growth yearly, and the number of publications increased slowly. The peak occurred in 2022, when articles jumped sharply to more than 30 publications. This was the most productive year during that period. But after that, in 2023 and 2024, the chart shows a fairly significant decline. The number of articles published has dropped drastically compared to the previous year.

From this pattern, there was an explosion of research activity in 2024, maybe because of greater support, increased funding, or a high research focus at that time. The decline in the following years could be due to reduced resources, a shift in research focus, or other external factors that influence it. Even so, in general, the trend that looks positive remains positive. The production of scientific articles tends to increase when viewed in the overall time, and this indicates good progress in research activities in the field. In addition to the annual publication trends, Figure 5 presents the distribution of average citations per published article each year, providing an overview of the scholarly impact of publications from 2016 to 2025.

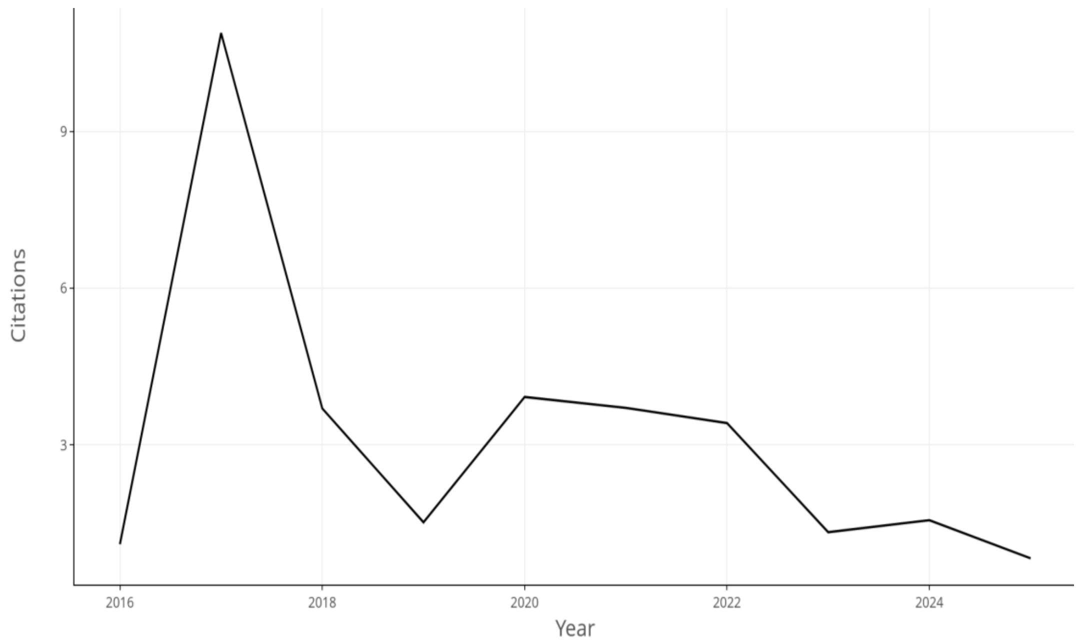


Figure 5. Annual Citation Chart

This can be explained in the form of Table 4.

Table 4. Annual Citation

Year	MeanTCperArt	N	MeanTCperYear	CitableYears
2016	11.00	1	1.10	10
2017	98.00	2	10.89	9
2018	29.60	5	3.70	8
2019	10.67	6	1.52	7
2020	23.50	10	3.92	6
2021	18.53	15	3.71	5
2022	13.70	23	3.42	4
2023	4.00	15	1.33	3
2024	3.13	38	1.56	2
2025	0.83	6	0.83	1

Figure 5 the graph and Table 4 shown provide an overview of the production of scientific articles each year, by looking at several important aspects, such as the average citations per article (MeanTCperArt), the number of articles published (N), the average citations per year (MeanTCperYear), and the number of years the article can be cited (CitableYears).

Only one article was published in 2016, but interestingly, it had a considerable impact, with an average of 11 citations. After that, the number of articles increased by two in 2017 and by five in 2018. However, although the number of articles increased, the average citations decreased, for example, to around 10.89 in 2017 and dropped further to 3.70 in 2018.

A jump in the number of articles was also seen in the following years, such as in 2020 and especially in 2021, which recorded 15 articles with an average of citations per year of around 3.71. But the most striking is from 2022 to 2024, when the number of articles rose sharply by 23 in 2022 and reached 38 in 2024. Unfortunately, this increase in number is not accompanied by a commensurate impact on citations. For example, in 2024, although many articles are published, the average citation will drop to 1.56 per article. The "publish or perish" mentality dominates academic culture, leading to a prioritization of quantity over quality in research output. This phenomenon is especially evident in fields like education, where researchers often chase trending topics instead of contributing meaningfully to societal needs, which is particularly pronounced in policy-driven fields such as education (Kara Aydemir & Can, 2019). The increasing pressure to publish rapidly can compromise the integrity and depth of scholarly work, as observed in various disciplines where authors gravitate towards quicker-to-produce article formats (Ngaage et al., 2023). Furthermore, in competitive academic environments, metrics such as impact factors and publication counts serve as benchmarks for success, compelling scholars to prioritize publication volume at the expense of thorough academic rigor (G. T. Lee et al., 2023). Consequently, this trend may contribute to a superficial approach to research, diminishing the overall quality and significance of scholarly contributions (S. Kim et al., 2018).

The year 2025 also shows a similar trend. Only six articles were published; the average citation was even lower, 0.83. The articles are probably still new and haven't been cited much. Overall, this data shows that although productivity in writing articles increases from year to year, not all of them have a high impact in terms of citations. To further illustrate the connections between authors, cited references, and research themes in this field, Figure 6 presents a Three-Field Plot diagram that maps the relationship among these three key elements in the analyzed literature.

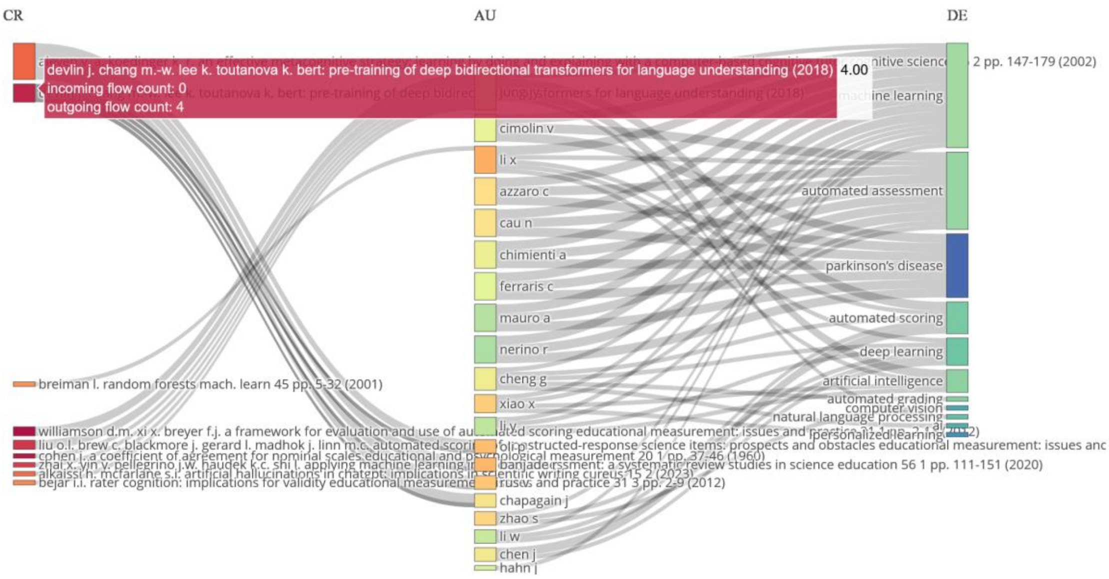


Figure 6. Three-Field Plot Diagram.

Figure 6 shows a Three-Field Plot diagram, which illustrates the relationship between three essential elements of research: frequently cited scientific references (CR), author names (AU), and research topics or themes discussed (DE). This diagram helps us see how the three are interconnected, who wrote what, and where they referred.

In the reference section (CR), two names stand out the most because they are most often cited: Aleven V. and Koedinger K. R., with their work on metacognitive strategies in computer-based learning, and Devlin J. and his team, famous for the BERT model, a preliminary training method for transformers in language processing. These two references are the primary references for many studies in this field.

If we look at the author section (AU), some names that often appear are Cimolin V., Li X., Azzaro C., and Chimenti A. They are known to write actively on machine learning, automatic assessment, and deep learning topics. This means they contribute a lot to research involving advanced technologies in education and data analysis.

Meanwhile, themes often appearing in the topic (DE) section include machine learning, automated assessment, deep learning, and artificial intelligence (AI). This shows that the authors are indeed focused on the use of modern technology in education, from automatic assessment to natural language processing using AI. The Three-Field Plot illustrates the relationship between authors, cited references, and research topics. It reveals how prominent authors cite key works, shaping the foundation of knowledge in AI in education. This connection shows how research on AI and automated assessment evolves and highlights topics like algorithmic bias in the literature. Conceptually, the diagram demonstrates how collaboration among authors and frequently cited references influences the direction of research in this field. The correlation between authors (AU), citations (CR), and research topics (DE) reflects how frequently cited authors shape the narrative on algorithmic bias. Major authors often focus on algorithmic bias in AI, which can exacerbate inequities in educational assessments. If left unaddressed, this bias could further perpetuate injustice and overlook more inclusive perspectives. Dominance of specific authors or topics can create scholarly hegemony, limiting the diversity of ideas in research. Over-reliance on a few authors or theories narrows the perspective on algorithmic bias, sidelining more equitable alternatives. Antonio Gramsci's theory of hegemony elucidates how intellectual dominance can lead to significant imbalances in knowledge production, marginalizing alternative perspectives. Gramsci argued that hegemony involves a combination of coercion and consent, where the ruling class's worldview is normalized and accepted as the societal standard, thereby stifling dissent and diverse viewpoints (Legwegoh & Fraser, 2015). This normalization often manifests in academia, where the paradigms prevalent in the Global North overshadow local knowledge systems from the Global South, perpetuating intellectual inequities. Gramsci delineates several forms of hegemony: total, decadent, and minimum, which illustrate the varying degrees of control exerted by dominant groups over societal discourse (Ramlan et al., 2023). Furthermore, the role of intellectuals in promoting hegemonic ideas is critical, as they help frame these dominant narratives as "common sense," thereby limiting the scope for alternative epistemologies (Richard & Molloy, 2020). The implications of Gramsci's insights are particularly relevant today, as the dominance of certain academic narratives continues to obscure broader perspectives in global knowledge production (Kharbach, 2020; James et al., 2022).

Overall, this diagram shows the close relationship between the authors, the references they use, and the topics they raise. They are all interconnected in one common thread: the application of innovative technology to support education's learning and evaluation process. The most relevant sources can be seen in Figure 7.

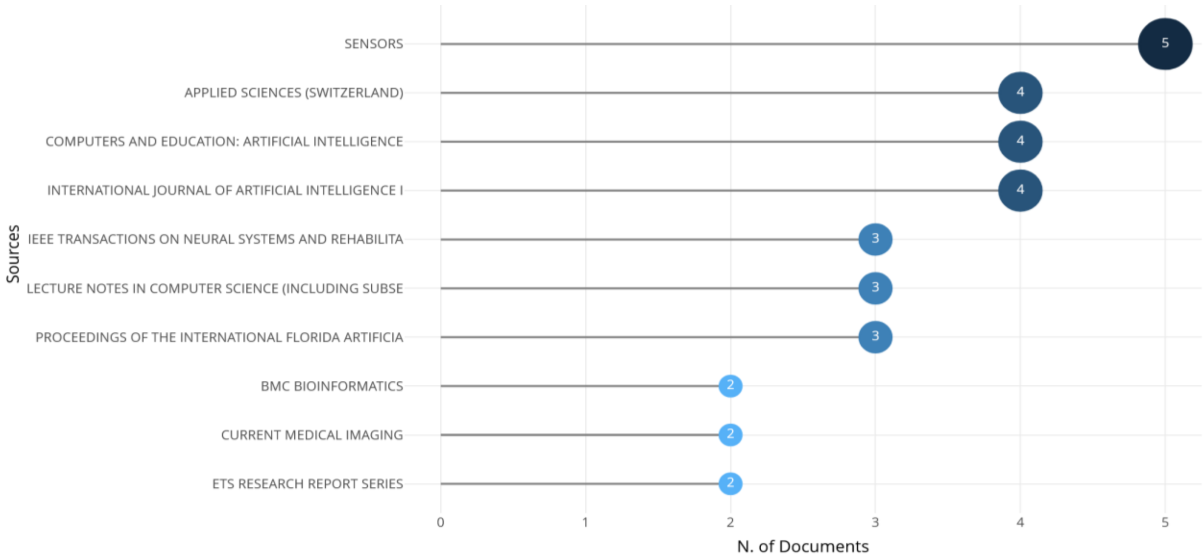


Figure 7. Most Relevant Source.

From the data shown in Figure 7, the journal SENSORS was recorded as the most published journal in this study, with five documents. This shows that the journal is vital in developing the topics discussed. The other three journals that are also quite prominent are Applied Sciences (Switzerland), Computers and Education: Artificial Intelligence, and the International Journal of Artificial Intelligence, each contributing four documents, which signifies their significant contributions in the related field.

In addition, several other journals, such as IEEE Transactions on Neural Systems and Rehabilitation Engineering, Lecture Notes in Computer Science (including subseries), and Proceedings of the International Florida Artificial Intelligence Research Society, publish three papers each. Although the number is slightly lower, these journals still have a fairly significant role.

Several other journals, such as BMC Bioinformatics, Current Medical Imaging, and ETS Research Report Series, contributed only two articles each. This shows that while they are still relevant, their contributions are not as significant as those of the major journals mentioned earlier.

Overall, Table 5 gives an idea of this study's most influential journals or sources of publications. The journal SENSORS and several other journals appear to dominate, which shows that the research topics are closely related to sensor technology, artificial intelligence in education, and computer science. However, the diversity of sources also indicates that this field of research touches a wide range of disciplines, from engineering and education to medicine and bioinformatics.

Table 5. List of Journals and Number of Articles Related to AI Research.

Number of Articles	Journal Name	Description
5	SENSORS	Published the highest number of articles in this study, highlighting its key role in advancing the discussed topics.
4	Applied Sciences (Switzerland), Computers and Education: Artificial Intelligence, International Journal of Artificial Intelligence	Following closely with significant contributions, underscoring their important influence in the field.
3	IEEE Transactions on Neural Systems and Rehabilitation Engineering, Lecture Notes in Computer Science (and its subseries), Proceedings of the International Florida Artificial Intelligence Research Society	While slightly lower in number, these journals maintain important roles in the research community.
2	BMC Bioinformatics, Current Medical Imaging, ETS Research Report Series	Their impact is smaller compared to the leading journals, but they remain relevant to the field.
Overall	Various sources from different disciplines	Illustrates the interdisciplinary nature of this research area, spanning engineering, education, medicine, and bioinformatics.

Table 6. Sources' Local Impact.

Source	h-index	g-index	m-index	TC	NP	PY_start
Sensors	4	5	0.800	160	5	2021
Ieee Transactions on Neural Systems and Rehabilitation Engineering	3	3	0.500	84	3	2020
International Journal of Artificial Intelligence in Education	3	4	0.600	58	4	2021
Bmc Bioinformatics	2	2	0.286	35	2	2019
Computers And Education: Artificial Intelligence	2	4	0.500	40	4	2022
Ets Research Report Series	2	2	0.250	9	2	2018
Journal Of Physics: Conference Series	2	2	0.250	44	2	2018
Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	2	3	0.333	36	3	2020
Sensors (Switzerland)	2	2	0.250	78	2	2018
2023 Ieee 10th International Conference on Data Science and Advanced Analytics, Dsaa 2023 - Proceedings	1	1	0.333	5	1	2023

Based on the data from Table 6, it can be seen that the SENSORS journal occupies the top position as the source with the most influence. The existing data reveals that the journals published after 2021 had a significant impact. Notably, the IEEE journals published after 2020 gained much attention with their high g and h indices. This demonstrates that since their publication, the content has been viewed and referred to by experts and scholars internationally. Known as one of the most dominant journals, the IEEE Transactions on Neural Systems and Rehabilitation Engineering demonstrates great influence in the field with an h-index of three, a g-index of three, an m-index of 0.50, and a total of 84 citations.

The International Journal of Artificial Intelligence in Education also has an h-index of 3, g-index 4, and m-index 0.6 created an impression proving their importance in the AI education sphere after 2021. Other less recognized journals like BMC Bioinformatics and Computers and Education: Artificial Intelligence, along with ETS Research, are still actively participating in the global dialogue despite having an h-index of two.

There are also journals or proceedings such as *Sensors (Switzerland)* and *the Journal of Physics: Conference Series,* which both have an h-index of 2, but with several publications and citations that are not as many as other top journals. The most recent source, the 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA 2023 – Proceedings), is recorded with an h-index and g-index of 1 each, an m-index of 0.333, and a total of 5 citations. It shows that the conference is still in its early stages of building its impact in the scientific community. Overall, this data shows a variation in the level of influence between journals. SENSORS and IEEE Transactions on Neural Systems and Rehabilitation Engineering are the most influential publications. In contrast, the others remain essential contributions in their respective fields, albeit on a smaller scale. Figure 8 is the Core Sources according to Bradford's Law data. This data is taken from the top 10 data.

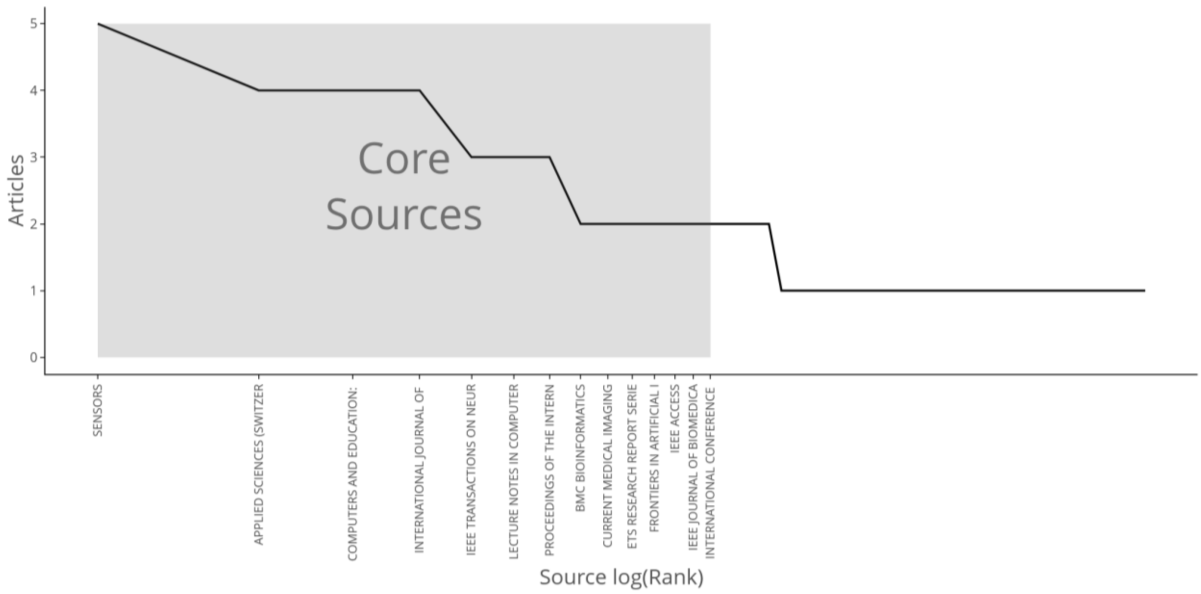


Figure 8. Core Sources by Bradford's Law.

Figure 8 shows the distribution of the primary sources of scientific publications under Bradford's Law. This graph shows how scientific articles are scattered across various journals or sources, following a pattern called Bradford's Law of Scattering. The horizontal axis depicts the order of sources based on logarithmic rank (log(Rank)), and the vertical axis shows the number of articles published by each source.

According to Bradford's principle, only a few core journals produce the majority of articles on a topic, while most other journals contribute only a small number of articles. This pattern is clearly visible on the chart. For example, the journal SENSORS is the most prominent because it publishes five articles, followed by several other journals such as Applied Sciences (Switzerland), Computers and Education, International Journal of Artificial Intelligence, and IEEE Transactions on Neural Systems and Rehabilitation Engineering, which publish four articles each.

The journals above fall into the "core sources" category marked on the graph with gray areas. After passing through this core group, the number of articles from subsequent sources decreases drastically, generally publishing only three or fewer articles. This graph shows that most scientific publications are concentrated in a few major journals, while the rest contribute only small amounts. This is based on Bradford's Law, which states that research activities in a field are usually focused on a few highly productive sources. Figure 9 is the Cumulative Occurrences graph data.

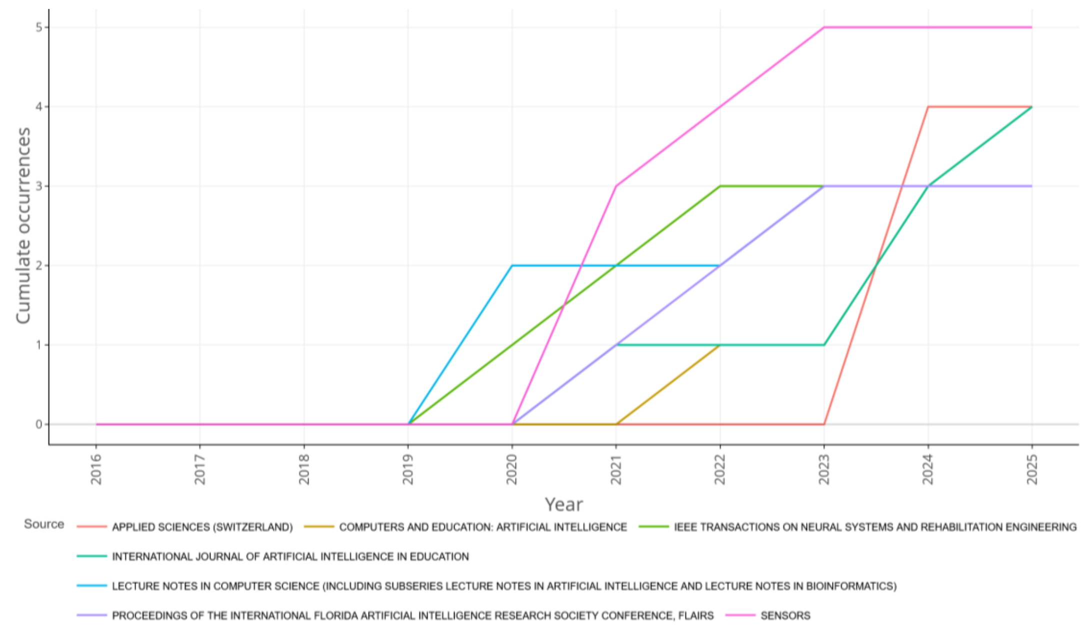


Figure 9. Cumulate Occurrences.

Figure 9 shows a cumulative graph of publications from various journals and conferences between 2016 and 2025. Each colored line on the graph represents a different source of publication, and shows how the number of articles from each source has increased over time. From the graph, it can be seen that the SENSORS journal (marked with a purple line) is the most prominent. Publications from this journal have been consistent since 2019 and experienced a sharp increase in 2023 and 2024, then peaked in 2025. It shows that SENSORS is increasingly being used as a reference and is becoming one of the primary sources in this field.

The IEEE Journal Transactions on Neural Systems and Rehabilitation Engineering (green) has also shown steady growth since its inception in 2020. Meanwhile, the International Journal of Artificial Intelligence in Education (in blue) began to see an increase in publications in the same year. However, the growth was not as fast as IEEE Transactions. Lecture Notes in Computer Science (turquoise) also became active in 2020, but its growth slowed after a few years, indicating that while it is still relevant, its contribution is not as significant as that of other sources. Meanwhile, the other two sources, Computers and Education: Artificial Intelligence (orange) and FLAIRS Conference Proceedings (yellow), show slower and less active publication growth in 2023 and 2024. This graph shows that SENSORS and IEEE Transactions are two of the most active and fast-growing sources in recent years, while other sources have grown more slowly or inconsistently.

Authors Information

Table 7. Most Relevant Authors

Authors	Articles	Articles Fractionalized
Li X	4	0.59
Banjade R	3	0.59
Oli P	3	0.59
Rus V	3	0.59
Xiao X	3	0.80
Zhao S	3	0.47
Azzaro C	2	0.21
Cau N	2	0.21
Chapagain J	2	0.34
Chen J	2	0.18

Table 7 shows the list of the most active authors in the publications analyzed. Two main pieces of information are shown: the total number of articles written by each author and *fractionalized contributions*, which measure how much of their role is in each article, especially when writing with other authors. From this data, it can be seen that LI X is the author with the highest number of articles, namely four articles, and has a fractionalization contribution of 0.59. This figure shows that his contribution to each article is still relatively significant despite working with other authors.

Authors such as BANJADE R, OLI P, RUS V, and XIAO X also wrote three articles each, with the same fractionalization value of 0.59. Meanwhile, ZHAO S also wrote three articles, but the fractionalized value was slightly lower at 0.47, indicating a collaborative role that may be more divided than other writers. Several authors, such as AZZARO C, CAU N, CHAPAGAIN J, and CHEN J, have written two articles each. However, their contributions per article vary between 0.18 and 0.34, which suggests that they are most likely to write alongside a larger team.

Overall, these data show that although some authors are quite productive in the number of articles they write, the level of their contributions in each article tends to be moderate or evenly split with other co-authors.

Data Authors Production Over Time.

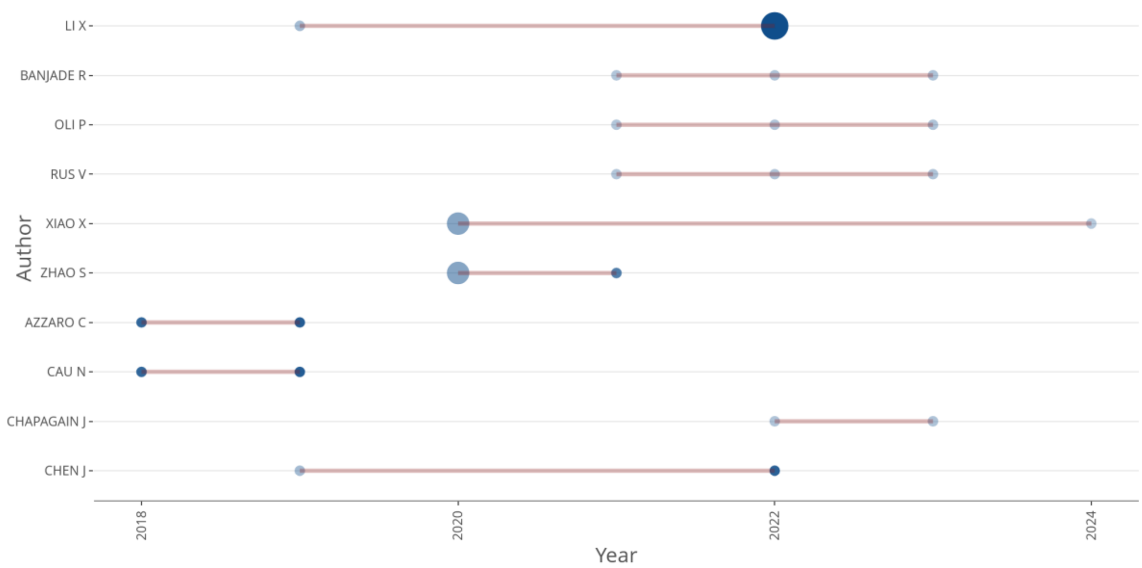


Figure 10. Data Authors Production Over Time

Figure 10 shows a graph of the productivity of writers from 2018 to 2024, showing how many articles they produce each year. Each line in the graph represents a single author, where the length and thickness of the lines reflect the number of articles and the publication period. This graph shows that LI X is the most prolific writer, with many articles, especially in 2020 and 2021, and continues to contribute until 2022. Meanwhile, authors such as BANJADE R, OLI P, and RUS V are also quite active, although the number of articles is not as high as LI X. They are consistent in publishing works, especially around 2020 and 2021.

Other authors such as XIAO X and ZHAO S are also actively involved, albeit with slightly fewer publications. Meanwhile, AZZARO C and CAU N have had fewer publications but remained active in publishing articles periodically throughout the period. On the other hand, the contributions of CHAPAGAIN J and CHEN J appear to be more limited. CHAPAGAIN J had several publications in 2020, while CHEN J was only recorded as having published one article in 2023.

Overall, this graph clearly shows who the most active authors in this field of research are. LI X seems to be the most prominent, while the other authors continue to play important roles despite the smaller number of publications.

Countries

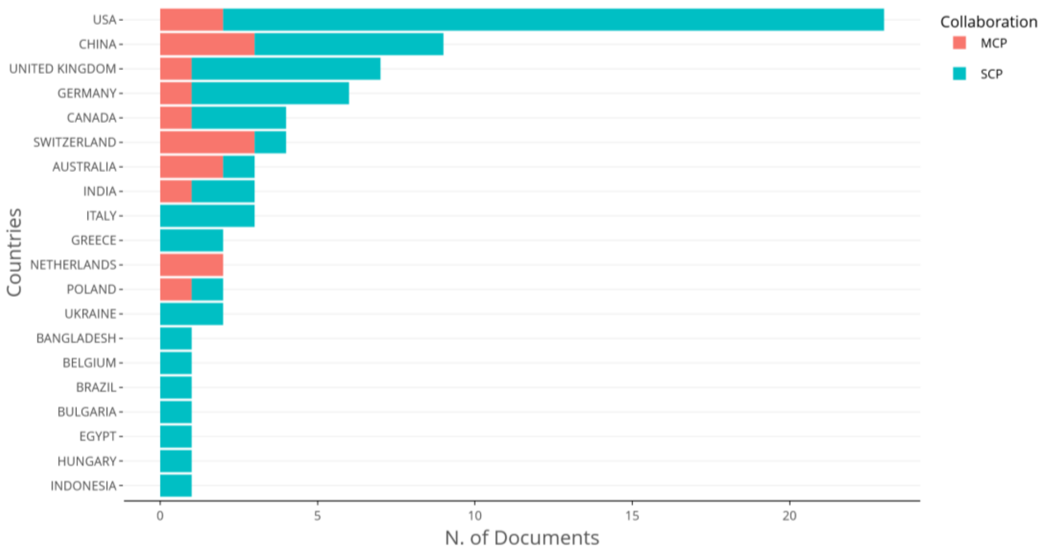


Figure 11. Author's Correspondence by Country

Figure 11 shows a graph of the number of scientific papers published by the country of origin of the author of the correspondence. This graph divides the types of collaboration in publications into two: SCP (Single-Country Papers), which means that authors are from only one country (marked in green), and MCP (Multi-Country Papers), which means that authors work together across countries (marked in red). The graph shows that the United States (USA) has the highest number of publications, and most of its articles are written by authors from the country (SCP). China also occupies the top position, with a reasonably balanced distribution between national publications and international collaborations.

Countries such as the United Kingdom and Germany follow this trend with a high number of publications, and most of them also come from domestic collaborations (SCPs). Then, countries such as Canada, Switzerland, Australia, and India also showed many publications, although most were still in the SCP category. Meanwhile, countries such as Italy, Greece, the Netherlands, and Poland have smaller contributions, but remain active, especially in domestic publications. Other countries such as Bangladesh, Belgium, Brazil, Bulgaria, Egypt, Hungary, and Indonesia were also recorded as contributing. However, the number of documents was smaller, and most were included in the SCP.

Global inequality in knowledge production illustrates significant disparities rooted in historical and structural factors. Researchers from the Global North tend to dominate academic landscapes, often shaping research agendas that reflect their cultural contexts and priorities. This dominance leads to an underrepresentation of voices from the Global South, where scholars face challenges such as language barriers and limited access to high-profile journals, which are predominantly published in English (Wang, 2022). Furthermore, it has been noted that theories developed in the Global North often inadequately address the realities of the Global South, contributing to inequalities in academic knowledge production (Hosford et al., 2022). Decolonial movements also emphasize that colonial legacies continue to influence knowledge production dynamics, highlighting the necessity for redistributive justice in research practices (Udah, 2024). Community-based participatory research (CBPR) represents a promising avenue for countering these inequities by democratizing knowledge creation, fostering collaboration, and empowering marginalized communities (Cohen & Snyder, 2024). Ultimately, addressing these disparities requires a collective commitment to reforming the academic landscape and promoting inclusive practices that elevate diverse perspectives in knowledge production (Wilcox et al., 2020)

This graph shows that large countries such as the USA and China dominate scientific publications in this field. Even so, there are also many contributions from various other countries, both through international cooperation and independent publications at the national level. Table 8 is the data of the top 10 Countries' Scientific Production.

Table 8. Top 10 Countries' Scientific Production.

Country	Freq
Usa	211
China	90
Uk	46
Germany	44
Italy	29
Switzerland	28
Canada	25
Netherlands	18
India	14
Japan	14

Table 8 shows the number of scientific publications produced by each country from 2016 to 2025. From the data, the United States (USA) is in the top position with 211 documents, showing how much it contributed to research during this period. China ranks second with 90 documents, quite large as well, although still far behind the USA. Next in order is the UK with 46 publications, followed by Germany, which recorded 44 documents, both of which continue to show a strong role in the development of global science. Italy and Switzerland are also on the list with 29 and 28 papers, which means they are active in their scientific contributions, although the scale is not as large as the countries above.

Canada and the Netherlands recorded 25 and 18 publications, respectively, while India and Japan followed with 14 documents per country. Although the number is smaller, these countries' contributions remain significant in the international research arena. Overall, this data shows a gap in the number of publications between countries, with the USA and China as major players, while other countries continue to play a role, albeit on a smaller scale.

Countries' Production over Time

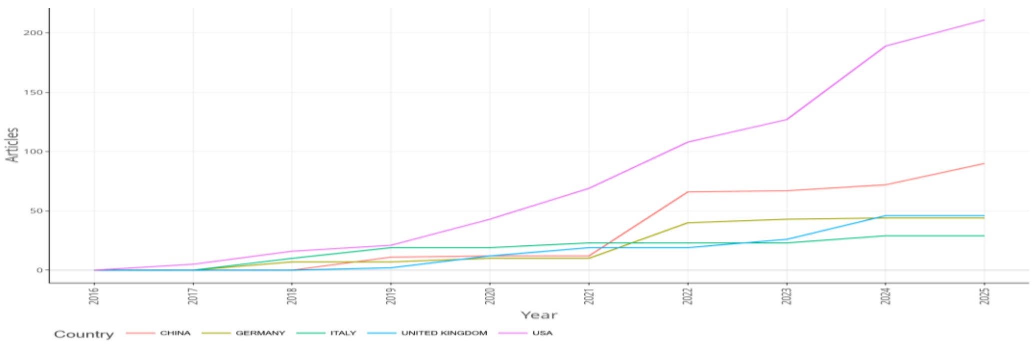


Figure 12. Charts of Countries' Production over Time

Figure 12 shows how scientific publication trends are developing in several countries, namely China, Germany, Italy, the United Kingdom (UK), and the United States (USA) from 2016 to 2025. In this graph, each line represents one country, with the horizontal axis indicating the year and the vertical axis indicating the number of articles published. The USA (with the purple line) looks the most prominent in the chart. After 2020, the number of publications from this country jumped sharply and continues to increase until 2025. It shows how dominant the USA's role is in the research world today. China (red line) also showed rapid growth that began to be felt since 2020, with a trend that continued to climb until the end of the period.

Meanwhile, Germany (green line) experienced a slower increase, but still showed steady growth. Italy (blue line) and the UK (yellow line) have a trend that tends to be flat with no significant spikes, but still increase consistently from year to year. Overall, this graph shows that the USA and China are major players in producing scientific articles, with very noticeable growth, especially in recent years. European countries such as Germany, Italy, and the UK are also actively contributing, albeit at a slower growth rate.

Most Cited Countries

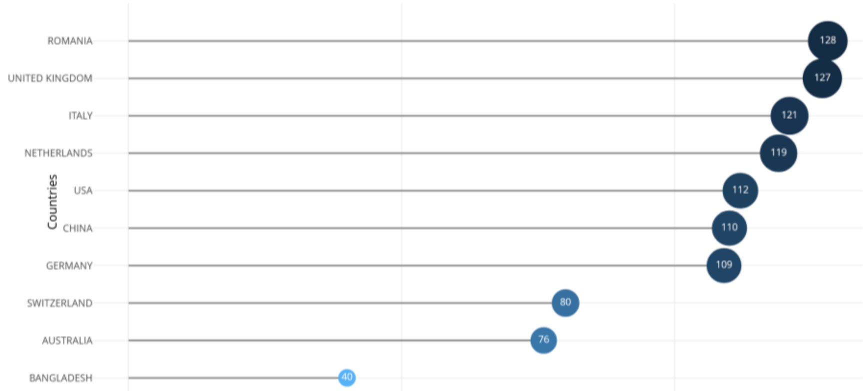


Figure 13. Most Cited Countries

Figure 13 shows a graph of the countries with the highest number of citations in the scientific publications analyzed. The vertical axis shows the country's name, while the horizontal axis shows the total citations received by publications from each country.

The graph shows that Romania is in first place with 128 citations, followed by the United Kingdom with 127 citations and Italy with 121 citations. It shows that these countries have a considerable impact on scientific publications. Furthermore, the Netherlands recorded 119 citations, while the USA and China received 112 and 110, respectively. Although the USA and China have more articles, they are not as dominant regarding the number of citations compared to other countries.

Countries such as Germany, Switzerland, Australia, and Bangladesh recorded lower citations. Germany received 80 citations, and Switzerland received 76 citations. Meanwhile, Bangladesh is in last place with only 40 citations, showing that despite contributing to scientific publications, its influence in terms of citations is still much smaller compared to other major countries.

Overall, this chart reveals that Romania, the United Kingdom, and Italy dominate in terms of citations. Although major countries like the USA and China have many publications, they are not so dominant in terms of citations.

Overlay visualization.

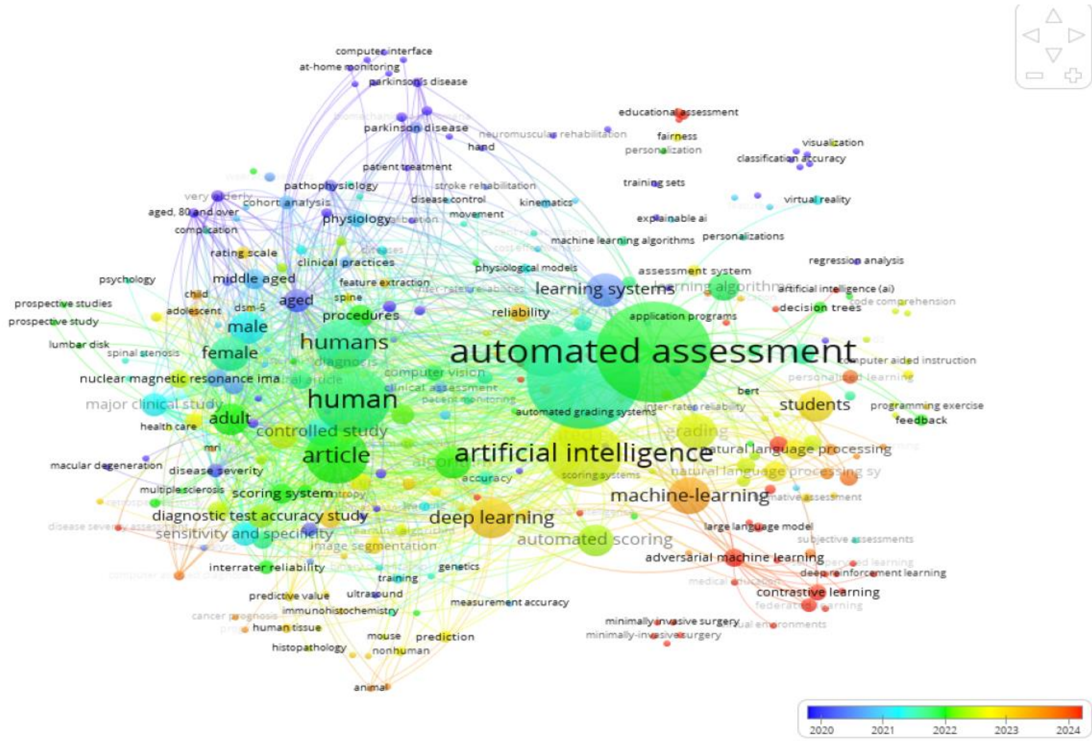


Figure 14. Contact Map

Figure 14 shows a keyword map showing the relationships between various topics related to "automated scoring," "artificial intelligence," and "machine learning" in scientific research published between 2020 and 2024. This map shows how many different ideas, technologies, and topics relate to each other.

It becomes clear that “automated assessment” and “artificial intelligence” are the two main focal points of the map because all the relevant keywords, such as “machine learning” and “deep learning,” circle them. It seems that in modern research, implementing AI and machine learning into automated assessments is a significant area of work. Some of the keywords like “student,” “educational assessment,” and “learning system”, suggest that a lot of research aims at leveraging AI for enhancing student assessments and learning.

Students are concerned about learning fairness and lack of transparency due to keywords like explainable AI and fairness. Some other topics that came up are “virtual reality,” “natural language processing,” and “predictive value,” which imply the introduction of novel technologies into AI and automated assessment. The emergence of “automated scoring systems” and “learning algorithms” is also indicative of the development of more sophisticated automated intelligent scoring systems. The evolving nature of these topics is also displayed in the map. Trends that are growing in popularity from 2020 to 2024 are represented by lighter colors. In recent times, there has been a greater emphasis on “contrastive learning”, “adversarial machine learning”, and “reinforcement learning”, which mark the application of sophisticated techniques in evaluative technology and machine learning.

Network Visualization

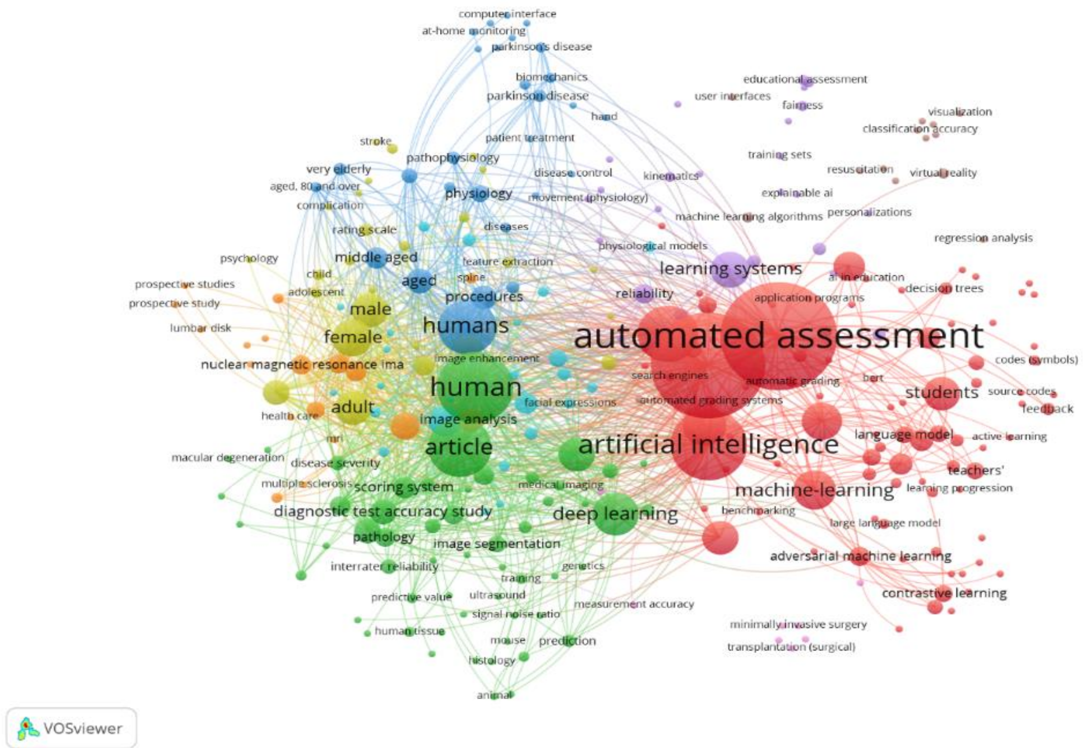


Figure 15. Network Visualization

Figure 15, a network visualization of the topics around automated assessment and artificial intelligence (AI) with a focus on its applications in education and medicine. The dots in the image represent key concepts one needs to understand, and the lines joining them indicate how they relate to each other. The largest grouping marked in red has a focus on automated assessments as well as AI. Keywords like AI, computerized assessment, machine learning, and deep learning appear in this group, which indicates that the core theme of this network revolves around the use of intelligent technology in assessments, especially supporting automated assessment systems in education and other relevant fields.

The yellow and green groups, however, are much more connected with medicine and human health, using keywords like spinal stenosis, human tissue, controlled studies, and diagnostic accuracy. It indicates that there is also an AI application in the diagnostic processes, as well as in other areas of human health research. Words such as student, educational assessments, and training sets connected to red groups show the importance of AI in education, particularly for assessing student progress and developing more personalized learning systems. Furthermore, the connection between AI, machine learning, and accuracy portrayed in this image shows the application of these technologies in enhancing precision in evaluations within education and healthcare.

As stated earlier, this AI-powered Image with an example displays the use of automated scoring in different fields. It also highlights the convergence of technology, aiming to achieve greater efficiency in automated systems as well as data-driven decision making.

SLR Study.

Of the 121 data points obtained, 82 will be focused on the type of article. To provide a clearer overview of the dominant research outcomes in the reviewed literature, Table 9 categorizes the main findings based on thematic focus, number of papers, and their corresponding percentages.

Table 9. Number of Documents by Publication Type.

Num	Kind	Sum
1	Article	82
2	Conference Paper	32
3	Reviews	7
	Total	121

Results of the PRISMA method

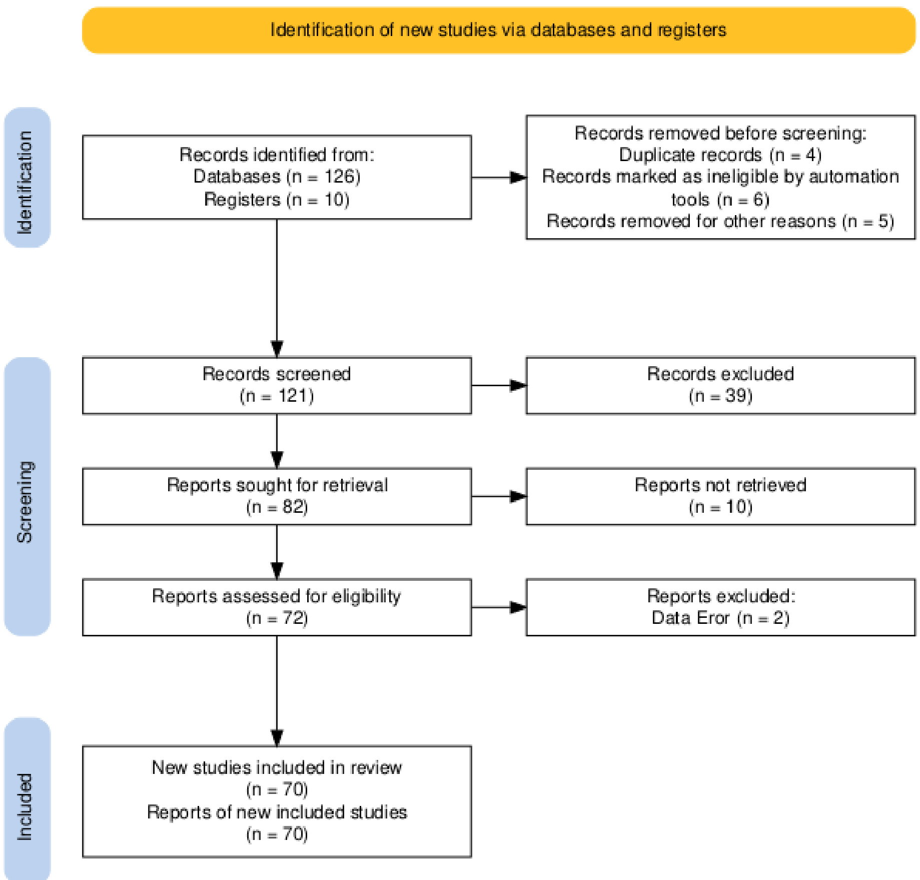


Figure 16. Results of the Prism Method.

Figure 16, the process of selecting articles for this systematic review is carried out according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and is described in detail through the flow chart in the PRISMA Figure above (Sohrabi et al., 2021). This process consists of four main stages, namely identification, screening, eligibility, and inclusion.

At the identification stage, the researcher obtained 136 records from two primary sources, namely from scientific databases, as many as 126, and from research records, as many as 10. Furthermore, early deletion of records that do not meet the basic criteria is carried out. A total of 15 records were issued, with details of 4 records being duplicates, six records being deemed unsuitable by the automatic selection tool, and five other records were deleted for different reasons that were not explicitly explained (e.g., records were incomplete or irrelevant to the topic of the study). Thus, the number of records that entered the screening stage was 121.

The screening stage involves examining the title and abstract of each note. This process aims to identify studies relevant to the review's focus. Of the 121 records, 39 were excluded because they did not meet the inclusion criteria, such as topic mismatches, irrelevant research methods, or inappropriate population coverage. The results of this stage resulted in 82 records that were deemed worthy of full-text search.

Furthermore, at the feasibility assessment stage, a search was carried out for the complete report of the 82 records. However, as many as 10 reports were not successfully accessed, so only 72 could be further assessed. Each report obtained is thoroughly analyzed to ensure methodological suitability, completeness of data, and relevance to the research question. Two reports were issued in this process due to data errors, such as inconsistencies between results and methods, or data that could not be processed systematically.

The last stage is inclusivity, where as many as 70 studies were declared to meet the criteria and included in the final review. This number is identical to the number of reports from the study, which means that each study has one main report analyzed. This selection process shows that out of 136 initial records, only about 51.5% managed to pass to the final stage, reflecting the application of a strict selection based on objective and structured criteria. Table 10 is a more precise classification of research methodologies based on the type of research (quantitative, experimental, observational, qualitative, conceptual, mixed, and systematic review), with the number of papers and the percentage.

Table 10. Classification of Types of Research Methods.

Types of Research	Number of Papers	Percentage	Characteristics & Examples
Quantitative	45	64.3%	- Using numerical data, statistical analysis, and computational models. - Examples: Research with machine learning, statistical analysis, or controlled experiments.
Experimental	15	21.4%	- Controlled design with intervention or treatment. - Example: Trial of AI models on specific datasets, laboratory experiments.
Observational	12	17.1%	- Observation without intervention, retrospective/prospective analysis. - Example: Clinical case study, secondary data analysis.
Qualitative	5	7.1%	- Interviews, content analysis, or phenomenological studies. - Examples: Studies of user perceptions of technology.
Mixed Methods	4	5.7%	- A combination of quantitative and qualitative approaches. - Examples: Surveys + interviews, or statistical analysis + case studies.
Systematic Review	6	8.6%	- Synthesis of literature with a structured methodology (PRISMA). - Example: Review of AI in education.
Conceptual	3	4.3%	- Theories, models, or frameworks without empirical data. - Example: A proposal for a new system architecture.
Total	**70**	**100%**	

Table 10 explains that quantitative research dominates (64.3%) due to trends in data, statistics, and algorithms such as machine learning. Experimental research (21.4%) is common in the medical and educational fields because it focuses on controlled trials, while observational research (17.1%) is more commonly used for retrospective studies without intervention. Qualitative research (7.1%) was used to explore perceptions or behaviors, and mixed methods (5.7%) combined numerical and narrative data to gain more comprehensive insights. A systematic review (8.6%) was conducted to synthesize the literature and identify trends, while conceptual research (4.3%) was theoretical with no empirical data. Overall, quantitative dominance reflects the trends of data-driven research and AI, while other approaches remain essential to social, pedagogical contexts, and theory development. Table 11 is the classification of the Main Findings based on the main findings category, the number of papers, and the percentage.

Table 11. Main Findings Classification.

Key Findings Categories	Number of Papers	Percentage	Specific Examples
Automated System Accuracy Improvement	28	40%	CNN achieved 98.62% accuracy in osteoarthritis classification (Paper 20)
Time/Process Efficiency	18	25.7%	AI reduces MRI reading time to 2.83 minutes (Paper 10)
Clinical Validation of New Tools	15	21.4%	SpineNet valid for LSS assessment (Paper 35)
Workload Reduction	12	17.1%	ML reduces assessment load by 64-74% (Paper 9)
Expert Equivalent Performance	10	14.3%	CVSnet is 600x faster with physician-equivalent accuracy (Paper 11)
Identify New Patterns/Correlations	8	11.4%	Negative correlation of task complexity with ML performance (Paper 14)
Cost Savings	5	7.1%	EyeArt® is $143 cheaper per patient (Paper 37)
Improved Learning Outcomes	5	7.1%	Students with AI feedback increased significantly (Paper 3)
Early/Predictive Detection	4	5.7%	Cardiovascular adequacy prediction model (Paper 68)

Table 11 explains that most studies (40%) show that AI/ML-based automated systems can improve the accuracy of assessments, often surpassing manual methods, especially in the classification of medical images and learning evaluation. In addition, efficiency was the main value raised, with a focus on reducing processing time (25.7%), workload (17.1%), and cost (7.1%). Clinical validation was also significant (21.4%), where many new tools were tested for conformity with medical standards. Some studies even uncover new patterns or correlations that are not detected manually and positively impact educational outcomes and skills. Looking ahead, research should focus more on integrating systems into real clinical practice, cost-benefit analysis of implementation, and longitudinal studies of impacts. Areas still underexplored include ethical issues, impact on the profession, and adaptation of cross-cultural and linguistic systems. Table 12 is the classification of design studies based on the study design category, the number of papers, and the percentage.

Table 12. Classification Study Design.

Categories Study Design	Number of Papers	Percentage	Characteristics & Examples
Observational Studies	32	45.7%	- Non-interventional data analysis- Examples: MRI retrospective assessment, clinical dataset analysis

Categories Study Design	Number of Papers	Percentage	Characteristics & Examples
Experimental Studies	18	25.7%	- Controlled design with intervention- Example: AI model trials, laboratory experiments
Computational Studies	12	17.1%	- Focus on algorithm/model development- Example: CNN training, ML architecture optimization
Systematic Review	6	8.6%	- Synthesis of structured literature- Example: Review of the application of AI in education
Qualitative Studies	5	7.1%	- Interviews, content analysis- Example: Study of user perception of technology
Case Studies	4	5.7%	- In-depth analysis of a case/specific- Example: System implementation in one institution
Conceptual Studies	3	4.3%	- Development of theories/frameworks- Example: New system architecture proposals

Table 12 shows the dominance of observational studies (45.7%), generally in retrospective analysis of medical data or existing datasets, such as MRI automatic assessment or student performance evaluation. Experimental studies (25.7%) were used to test the effectiveness of new models or technologies in a controlled environment, while computational studies (17.1%) focused on algorithm development and machine learning architecture optimization. Non-empirical studies are also present in systematic reviews (8.6%) to synthesize existing literature and conceptual studies (4.3%) that propose new theories or frameworks. Future research is suggested to increase longitudinal studies to assess long-term impacts, comparative studies between methods, and real implementation in the field. In addition, there is a need to strengthen the use of mixed study designs, replication studies to ensure the validity of findings, and larger, more diverse samples. Table 13 is the classification of future research based on the recommendation category, the number of papers, and the percentage:

Table 13. Future Research Classification.

Recommended Categories	Number of Papers	Percentage	Specific Examples
Model/Algorithm Improvement	38	54.3%	- Development of more accurate architectures (Paper 44)- Optimization of model parameters (Paper 52)
Clinical/Field Validation	29	41.4%	- Multicenter trial (Paper 10)- Longitudinal study (Paper 35)
Dataset Expansion	25	35.7%	- Increase in the number of samples (Paper 19)- Diversification of the population (Paper 37)
Technology Integration	22	31.4%	- Multi-modal combination (Paper 18)- IoT and edge computing (Paper 6)
Clinical/Practical Applications	18	25.7%	- Implementation in hospitals (Paper 11)- Physician decision aids (Paper 36)
Ethics & Security Aspects	12	17.1%	- Algorithmic bias handling (Paper 56)- Patient data privacy (Paper 64)
Cost Optimization & Scalability	10	14.3%	- Reduced implementation costs (Paper 5)- Affordable systems (Paper 65)
Comparative Studies	8	11.4%	- Comparison between methods (Paper 20)- Performance benchmarking (Paper 33)
Standard Development	6	8.6%	- Implementation guidelines (Paper 15)- Standard evaluation criteria (Paper 48)

Table 13 shows that most studies (54.3%) focus on improving models or algorithms, while 31.4% recommend integration with other supporting technologies. A total of 41.4% highlighted the importance of more rigorous validation, and 25.7% encouraged practical implementation in the real world. Interestingly, non-technical aspects, such as ethical and safety issues (17.1%) and cost and scalability considerations (14.3%), are starting to receive attention. Therefore, future research should prioritize efforts to bridge technology development with real implementation, consider ethics and regulations from the design stage, and conduct replication studies and method comparisons. Underexplored areas include the social impact of automated technologies, systems sustainability analysis, and cross-cultural and language adaptations that ensure relevance and wide acceptance. Table 14 is the classification of the Summary of the discussion based on the discussion category, the number of papers, and the percentage.

Table 14. Classification Summary of the Discussion.

Discussion Categories	Number of Papers	Percentage	Representative Examples
System Performance Validation	42	60%	"Model shows expert-equivalent accuracy with improved time efficiency" (Paper 10)
Practical Implications	35	50%	"Findings support clinical implementation as a decision aid" (Paper 11)
Study Limitations	31	44.3%	"Key limitations on sample size and dataset variation" (Paper 18)
Potential for Development	28	40%	"Integration with PACS can improve clinical adoption" (Paper 10)
Comparison with Previous Studies	22	31.4%	"Results are consistent with study X but differ in aspect Y" (Paper 25)
Social/Technical Impact	15	21.4%	"Systems can reduce disparities in access to health services" (Paper 37)
Policy Recommendations	9	12.9%	"Validation standards are required for clinical implementation" (Paper 29)
Theoretical Implications	7	10%	"Findings support Z's theory of neural adaptation" (Paper 55)

Table 14 shows that most studies (60%) focus on validating system performance through quantitative metrics such as accuracy, sensitivity, and time efficiency. In addition, 50% of the discussion emphasized practical applications and factors influencing technology adoption in real institutions. Transparency on the study's limitations was also quite dominant (44.3%), generally related to sample issues, data bias, or generalization limitations. Balanced discussions are also emerging, with 40% including development directions and 31.4% conducting literature comparisons. To improve the quality of the discussion, it is suggested that future research discuss more socio-technical impacts, relate findings to relevant theories, and include more in-depth causal analysis. Areas that still need to be strengthened include the discussion of implementation ethics (only 12.9%), cost-benefit comparative analysis, and the long-term impact of the use of automated systems.

The findings of this article align with and expand upon previous research that highlights both the promising benefits and significant challenges of AI in academic assessment. Similar to studies by Luckin et al. (2022) and Crompton & Burke (2023), this review confirms that AI enhances efficiency, consistency, and personalization in grading processes, thereby improving educational outcomes. However, consistent with Gándara et al. (2024), it also underscores persistent issues of algorithmic bias stemming from unrepresentative training data and socio-cultural influences. Unlike some earlier works that primarily focused on technical performance, this research offers a more comprehensive analysis by integrating bibliometric data and emphasizing the ethical dimensions and the necessity for educator training and transparent guidelines. This holistic perspective strengthens the argument for responsible AI deployment in education by balancing technological advances with fairness considerations.

The impact of AI on benefit assessments in education is profound, offering scalable solutions that can reduce educator workload, provide timely and standardized feedback, and support personalized learning pathways. These advantages have the potential to transform traditional assessment methods, making them more adaptive to individual student needs. Nevertheless, mitigation strategies for algorithmic bias are crucial to safeguard against unintended discrimination. Effective approaches include diversifying training datasets to better represent all student populations, implementing transparent and explainable AI models, and involving educators in designing and overseeing AI assessment tools. Continuous monitoring and ethical guidelines are also vital to ensure that AI supports equitable learning environments rather than exacerbating existing inequalities. Together, these strategies can help maximize the benefits of AI while minimizing its risks in academic assessment.

Implication

This research highlights the significant potential of artificial intelligence (AI) to transform educational assessment by enhancing efficiency, consistency, and personalization. The findings imply that AI can substantially reduce educators' workload by automating grading processes and delivering timely, standardized student feedback, leading to improved learning outcomes. Additionally, the research stresses the critical importance of addressing algorithmic bias to ensure fairness and equity in educational settings. The study advocates for responsible AI deployment that balances technological advancements with social justice by emphasizing ethical guidelines, transparency in AI models, and educator training. These implications provide a roadmap for policymakers, developers, and educators to collaborate in creating AI-driven assessment tools that are both effective and inclusive.

LIMITATION

However, this study also faces several limitations that should be acknowledged. The analysis is restricted to articles published between 2021 and 2025 and sourced solely from the Scopus database, which may exclude relevant research from other periods or databases, potentially narrowing the comprehensiveness of the review. Furthermore, while the systematic literature review and bibliometric analysis provide a broad overview, they cannot fully capture AI implementation's nuanced, real-world complexities in diverse educational contexts. The dynamic and rapidly evolving nature of AI technologies means that findings may quickly become outdated as new algorithms and ethical frameworks emerge. Future research would benefit from expanding temporal and database scope and incorporating empirical studies that examine the long-term effects and practical challenges of AI-based assessments in varied cultural and institutional environments.

CONCLUSION

This research provides a comprehensive overview of the role of AI in educational assessment systems, highlighting both its benefits and challenges. On the positive side, AI has demonstrated significant potential to enhance assessment efficiency, improve consistency, and enable personalized learning experiences that enrich student outcomes. Automation reduces educators' workload, allowing faster and more objective feedback delivery. However, alongside these benefits lie substantial challenges, particularly algorithmic bias arising from unrepresentative training data, developer prejudices, and socio-cultural influences. These biases risk perpetuating inequities, especially for marginalized student groups.

Furthermore, AI's limitations in capturing the unique context of each learner and the potential overreliance on automated systems add complexity to its practical adoption. Consequently, this research underscores the necessity for a balanced and responsible approach to AI deployment in education. Clear ethical guidelines, algorithmic transparency, and comprehensive educator training are essential to mitigate bias and uphold fairness in assessment processes. Future research should focus on developing inclusive and culturally adaptive AI models, alongside rigorous validation and long-term evaluation of their real-world impact. By doing so, the advantages of AI can be maximized without compromising principles of equity and justice, ensuring that AI serves as an effective and fair educational tool for all learners.

AUTHOR CONTRIBUTIONS

The contributions of each author to this article are as follows: The conceptualization was carried out by ER, RAR, and AZAR. The methodology was developed by ER. The validation process was conducted by ER. Formal analysis and investigation were carried out by ER, RAR, and AZAR, who also contributed to the provision of resources and data curation. The original draft preparation was written by ER, with writing, proofreading, review, and editing performed by ER, HAK, and JF.

REFERENCES

Aghaziarati, A., Nejatifar, S., & Abedi, A. (2023). Artificial intelligence in education: Investigating teacher attitudes. *AI and Tech in Behavioral and Social Sciences*, *1*(1), 35–42. https://doi.org/10.61838/kman.aitech.1.1.6

Anuyahong, B., Rattanapong, C., & Patcha, I. (2023). Analyzing the impact of artificial intelligence in personalized learning and adaptive assessment in higher education. *International Journal of Research and Scientific Innovation*, *10*(4), 88–93. https://doi.org/10.51244/IJRSI.2023.10412

Armoogum, S., & Zakaria, M. Z. (2024). Deep Learning innovations in fingerprint recognition : A comparative study of model efficiencies. *International Journal of Advances in Artificial Intelligence and Machine Learning Vol*, *1*(1), 28–35. https://doi.org/10.58723/ijaaiml.v1i1.294

Baker, C., & Fairclough, S. H. (2021). Adaptive virtual reality. *Current Research in Neuroadaptive Technology*, 1(1),159–176. https://doi.org/10.1016/B978-0-12-821413-8.00014-2

Branch, R. M. (2009). *Instructional design-the ADDIE approach.* Springer. https://doi.org/10.1007/978-0-387-09506-6

Cavique, L. (2024). Implications of causality in artificial intelligence. *Frontiers in Artificial Intelligence*, *7*(August), 1–5. https://doi.org/10.3389/frai.2024.1439702

Coghlan, S., Miller, T., & Paterson, J. (2021). Good proctor or "big brother"? Ethics of online exam supervision technologies. *Philosophy and Technology*, *34*(4), 1581–1606. https://doi.org/10.1007/s13347-021-00476-1

Cohen, A. K., & Snyder, R. E. (2024). Community-based participatory research for epidemiology, health equity, and community goals: Insights from Brazil, France, and USA. *Community Health Equity Research and Policy*. 45(4):385-393. https://doi.org/10.1177/2752535X241262857

Crompton, H., & Burke, D. (2022). Artificial intelligence in K-12 education. *SN Social Sciences*, *2*(7), 2(3), 431-440. https://doi.org/10.1007/s43545-022-00425-5

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education : The state of the field. *International Journal of Educational Technology in Higher Education*, *20*(22), 1–22. https://doi.org/10.1186/s41239-023-00392-8

Damaševičius, R. (2024). Commentary: ChatGPT-supported student assessment – can we rely on it? *Journal of Research in Innovative Teaching and Learning*, *17*(2), 414–416. https://doi.org/10.1108/JRIT-09-2024-195

Emilio, F. (2024). FAIRNESS and bias in artificial intelligence : A brief survey of sources , impacts , and mitigation strategies. *Sci*, *6*(1), 1–18. https://doi.org/10.3390/sci6010003

Gándara, D., Anahideh, H., Ison, M. P., & Picchiarini, L. (2024). Inside the black box: Detecting and mitigating algorithmic bias across racialized groups in college student-success prediction. *AERA Open*, *10(1)*. https://doi.org/10.1177/23328584241258741

Herawati, A. A., Yusuf, S., Ilfiandra, I., Taufik, A., & Ya Habibi, A. S. (2024). Exploring the role of artificial intelligence in education, students preferences and perceptions. *Al-Ishlah: Jurnal Pendidikan*, *16*(2), 1029–1040. https://doi.org/10.35445/alishlah.v16i2.4784

Hosford, K., Aqil, N., Windle, J., Gundur, R. V., & Allum, F. (2022). Who researches organised crime? A review of organised crime authorship trends (2004–2019). *Trends in Organized Crime*, *25*(3), 249–271. https://doi.org/10.1007/s12117-021-09437-8

Ifenthaler, D., Majumdar, R., Gorissen, P., Judge, M., Mishra, S., Raffaghelli, J., & Shimada, A. (2024).

Artificial intelligence in education: Implications for policymakers, researchers, and practitioners. *Technology, Knowledge and Learning*, *29*(4), 1693–1710. https://doi.org/10.1007/s10758-024-09747-0

James, K. E., Tuidraki, H. D., & Tanzil, S. A. (2022). Postcolonial control of Fiji soccer and the return of subjugated knowledges: From the 1970s to the 2010s. *Frontiers in Sports and Active Living*, *4(1), 1-5*. https://doi.org/10.3389/fspor.2022.1005733

Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability (Switzerland)*, *15*(16), 1–27. https://doi.org/10.3390/su151612451

Kara Aydemir, A. G., & Can, G. (2019). Educational technology research trends in Turkey from a critical perspective: An analysis of postgraduate theses. *British Journal of Educational Technology*, *50*(3), 1087–1103. https://doi.org/10.1111/bjet.12780

Kasim, S., Zakaria, M. Z., & Efrizoni, L. (2025). AI and the optimization of product placement : Enhancing sales through strategic positioning. *International Journal of Advances in Artificial Intelligence and Machine Learning Vol*, *2*(1), 18–26. https://doi.org/10.58723/ijaaiml.v2i1.381

Kharbach, M. (2020). Understanding the ideological construction of the Gulf crisis in Arab media discourse: A critical discourse analytic study of the headlines of Al Arabiya English and Al Jazeera English. *Discourse and Communication*, *14*(5), 447–465. https://doi.org/10.1177/1750481320917576

Khater, A. S., Zaaqoq, A. A., Wahdan, M. M., & Ashry, S. (2023). Knowledge and attitude of Ain Shams University Medical Students towards artificial intelligence and its application in medical education and practice. *Educational Research and Innovation Journal*, *3*(10), 29–42. https://doi.org/10.21608/erji.2023.306718

Khosravi, H., Sadiq, S., & Amer-Yahia, S. (2023). Data management of AI-powered education technologies: Challenges and opportunities. *Journal of Learning Letters*, *1(1)*, 1–11. https://doi.org/10.59453/XLUD7002

Kim, F., Williams, L. A., Johnston, E. L., & Fan, Y. (2024). Bias intervention messaging in student evaluations of teaching: The role of gendered perceptions of bias. *Heliyon*, *10*(17), 37140. https://doi.org/10.1016/j.heliyon.2024.e37140

Kim, S., Chung, E., & Lee, J. Y. (2018). Latest trends in innovative global scholarly journal publication and distribution platforms. *Science Editing*, *5*(2), 100–112. https://doi.org/10.6087/kcse.133

Lee, G. T., Jiang, Y., & Hu, X. (2023). Brief report: Publications from mainland China, Hong Kong, and Taiwan in behavioral journals 1980–2021. *Behavioral Interventions*, *38*(3), 793–803. https://doi.org/10.1002/bin.1947

Lee, J., Hicke, Y., Yu, R., Brooks, C., & Kizilcec, R. F. (2024). The life cycle of large language models in education: A framework for understanding sources of bias. *British Journal of Educational Technology*, *55*(5), 1982–2002. https://doi.org/10.1111/bjet.13505

Legwegoh, A. F., & Fraser, E. D. G. (2015). Food crisis or chronic poverty: Metanarratives of food insecurity in sub-saharan africa. *Journal of Hunger and Environmental Nutrition*, *10*(3), 313–342. https://doi.org/10.1080/19320248.2014.962777

Luckin, R., Cukurova, M., & Kent, C. (2022). Computers and education : Artificial intelligence empowering educators to be ai-ready. *computers and education: Artificial Intelligence*, *3(1)*, 100076. https://doi.org/10.1016/j.caeai.2022.100076

Luckin, R., Cukurova, M., Kent, C., & du Boulay, B. (2022). Empowering educators to be AI-ready. *Computers and Education: Artificial Intelligence*, *3(1)*, 100076. https://doi.org/10.1016/j.caeai.2022.100076

Ngaage, L. M., Borrelli, M. R., Ketheeswaran, S., & Shores, J. T. (2023). Article factors influencing gender disparities in senior authorship of plastic surgery publications. *Annals of Plastic Surgery*, *91*(6), 638–643. https://doi.org/10.1097/SAP.0000000000003709

Opesemowo, O. A. G., & Adekomaya, V. (2024). Harnessing artificial intelligence for advancing sustainable development goals in south africa's higher education system: A qualitative study. *International Journal of Learning, Teaching and Educational Research*, *23*(3), 67–86. https://doi.org/10.26803/ijlter.23.3.4

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(8), 2307. https://doi.org/10.29333/ejmste/13428

Pham, N., Pham Ngoc, H., & Nguyen-Duc, A. (2025). Fairness for machine learning software in education: A systematic mapping study. *Journal of Systems and Software*, *219*(1), 112244. https://doi.org/10.1016/j.jss.2024.112244

Pranckutė, R. (2021). Scopus and Web of Science stands out for systematic reviews, offering comprehensive coverage across disciplines, including journals, conferences, and patents. *Publications*, *9*(1), 1–59. https://doi.org/10.3390/publications9010012

Ramadhan, M. R. (2024). Pengaruh konsep keadilan dalam Al Qur'an dan relevansinya dalam hukum manusia, *Journal of International Multidisciplinary Research, 2(11),* 134-136. https://doi.org/10.62504/jimr972

Ramlan, H., Kasnawi, T., Manda, D., Kamaruddin, S., Syukur, M., & Suardi. (2023). Hierarchy of child exploitation by parents in Makassar city, Indonesia. *International Journal of Arts and Humanities Studies*, *3*(4), 42–47. https://doi.org/10.32996/ijahs.2023.3.4.6

Richard, K., & Molloy, S. (2020). An Examination of emerging adult military men: Masculinity and U.S. military climate. *Psychology of Men and Masculinity*, *21*(4), 686–698. https://doi.org/10.1037/men0000303

Sasilatha, T., & Suprianto, A. A. (2025). AI-driven approaches to power grid management : Achieving efficiency and reliability. *International Journal of Advances in Artificial Intelligence and Machine Learning Vol*, *2*(1), 27–37. https://doi.org/10.58723/ijaaiml.v2i1.380

Sohrabi, C., Franchi, T., Mathew, G., Kerwan, A., Nicola, M., Griffin, M., Agha, M., & Agha, R. (2021). PRISMA 2020 statement: What's new and the importance of reporting guidelines. *International Journal of Surgery*, *88*(2), 39–42. https://doi.org/10.1016/j.ijsu.2021.105918

Tri, M. F., & Nataliani, Y. (2021). Analisis pengaruh penilaian asesor terhadap kinerja guru mata pelajaran dengan k-means clustering. *Indonesian Journal of Computing and Modeling*, *4*(1), 14–22. https://doi.org/10.24246/icm.v4i1.5063

Udah, H. (2024). Decolonising research for justice: Ethical imperatives and practical applications. *International Journal of Qualitative Methods , 23(1),* 2-4. https://doi.org/10.1177/16094069241294040

Ullah, R., Asghar, I., & Griffiths, M. G. (2023). An integrated methodology for bibliometric analysis: A case study of internet of things in healthcare applications. *Sensors*, *23*(1), 67. https://doi.org/10.3390/s23010067

Vieriu, A. M., & Petrea, G. (2025). The impact of Artificial Intelligence (AI) on students' academic development. *Education Sciences*, *15*(3), 1–12. https://doi.org/10.3390/educsci15030343

Wang, Q. (2022). "The very interesting finding suggests that": A cognitive frame-based analysis of interest markers by authors' geo-academic location in applied linguistics research articles. *Frontiers in Psychology*, *13(1),* 2-5. https://doi.org/10.3389/fpsyg.2022.1020854

Wilcox, E. S., Chimedza, I. T., Mabhele, S., Romao, P., Spiegel, J. M., Zungu, M., & Yassi, A. (2020). Empowering health workers to protect their own health: A study of enabling factors and barriers to implementing healthwise in Mozambique, South Africa, and Zimbabwe. *International Journal of Environmental Research and Public Health*, *17*(12), 1–17. https://doi.org/10.3390/ijerph17124519

Yakir, J., Valentin, O., & Rosario, T. (2023). Bibliometric analysis of youtube platform research trends using the vosviewer application. *Journal Emerging Technologies in Education*, *1*(1), 69–81. https://doi.org/10.55849/jete.v1i1.175

Zong, R., Zhang, Y., Stinar, F., Shang, L., Zeng, H., Bosch, N., & Wang, D. (2023). A crowd–AI collaborative approach to address demographic bias for student performance prediction in online education. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *11*(1), 198–210. https://doi.org/10.1609/hcomp.v11i1.27560