An-Najah University Journal for Research – B
Humanities
مجلة جامعة النجاح للأبحاث – ب
العلوم الإنسانية

A Comprehensive Bibliometric and Systematic Literature Review on AI-Assisted Assessment Tools in Education

Eko Risdianto[1,*], Rada Kristina[1], Deshinta Arrova Dewi[2], Ahmad Zabidi Abdul Razak[3], M. Abdul Jamal[4], Adamu Abubakar Muhammad[5] & Hera Anis Kartika[6]

(Type: Full Article). Received: 26th Jul. 2025, Accepted: 10th Nov. 2025, Published: ××××. DOI: https://doi.org/10.xxxx.

Accepted Manuscript (In Press)

Abstract: Objectives: This study aims to map global research trends on AI-assisted assessment tools, evaluate their effectiveness in improving educational evaluation, and identify challenges that influence their adoption across learning environments. **Methodology:** A combined bibliometric and systematic literature review (SLR) approach was used. A total of 218 Scopus-indexed publications (2015–2025) were collected using three keyword search strategies. After duplicate removal and PRISMA-based screening, 86 studies were included for qualitative synthesis. Bibliometric indicators—such as publication growth, citation patterns, keyword co-occurrence, and country contributions were analyzed using Biblioshiny. The SLR examined methodological classifications, assessment outcomes, user acceptance, and ethical considerations. **Result:** The result show rapid growth in publications with an annual increase of 31.95%, dominated by themes such as ChatGPT, machine learning, and learning analytics. The United States leads global scientific production with 397 publications, followed by the United Kingdom and China. AI-assisted assessment tools improve grading consistency, reduce human bias, support automated feedback, and enhance assessment efficiency. However, adoption is uneven due to infrastructural limitations, data-related constraints, algorithmic transparency issues, and educators' concerns regarding fairness and autonomy. Most studies focused on academic performance (32.9%) and health-related applications (22.4%), while ethical and inclusivity aspects remain underexplored. **Conclusions:** AI assisted assessment tools have significant potential to transform evaluation practices by enhancing fairness, accuracy, and personalized learning support. Nevertheless, effective implementation requires addressing institutional readiness, improving educator competence, and ensuring transparent decision-making processes to maintain trust and accountability. **Recommendations:** Future research should focus on long-term classroom evaluations, development of culturally responsive AI models, and broader investigation of ethical risks such as bias and data privacy. Educational institutions and policymakers are encouraged to integrate AI-guided assessment frameworks with human-led pedagogical judgment, ensuring equitable and responsible use of AI technologies in education.

Keywords: Artificial Intelligence, AI-Assisted Assessment, Educational Technology, Personalized Feedback, Assessment Bias, Educational Innovation, AI In Education.

مراجعة شاملة للكتابات المرجعية والمنهجية حول أدوات التقييم المدعومة بالذكاء الاصطناعي في مجال التعليم

إيكو ريسديانتو[1،*]، و رادا كريستينا[1]، وديشينتا أروفا ديوي[2]، وأحمد زابيدي عبد الرزاق[3]، وم. عبد الجمال[4]، و آدامو أبوبكر محمد[5]، وهيرا أنيس كرتيكا[6]

تاريخ التسليم: (2025/7/26)، تاريخ القبول: (2025/11/10)، تاريخ النشر: (××××)

الملخص: الأهداف: تهدف هذه الدراسة إلى رسم خريطة اتجاهات البحث العالمية حول أدوات التقييم المدعومة من الذكاء الاصطناعي، وتقييم فعاليتها في تحسين التقييم التعليمي، وتحديد التحديات التي تؤثر على تبنيها عبر بيئات التعلم. **المنهجية:** تم استخدام نهج مشترك بين الببليومترية ومراجعة الأدبيات المنهجية (SLR). تم جمع ما مجموعه 218 منشورا مفهرسا ب Scopus (2015–2025) باستخدام ثلاث استراتيجيات بحث بالكلمات المفتاحية. بعد إزالة النسخ والفحص المعتمد على PRISMA، تم تضمين 86 دراسة للتخليق النوعي. تم تحليل مؤشرات البليومترية — مثل نمو النشر، أنماط الاستشهاد، تداخل الكلمات المفتاحية، ومساهمات الدول باستخدام Biblioshiny. فحص SLR التصنيفات المنهجية، ونتائج التقييم، وقبول المستخدمين، والاعتبارات الأخلاقية. **النتيجة:** تظهر النتيجة نموا سريعا في المنشورات بزيادة سنوية بنسبة 31.95٪، تهيمن عليها مواضيع مثل ChatGPT، وتعلم الآلة، وتحليلات التعلم. تتصدر الولايات المتحدة الإنتاج العلمي العالمي ب 397 منشورا، تليها المملكة المتحدة والصين. تعزز أدوات التقييم المدعومة الذكاء الاصطناعي من اتساق التقييم، وتقلل من التحيز البشري، وتدعم التغذية الراجعة المؤتمتة، وتعزز كفاءة التقييم. ومع ذلك، فإن التبني غير متساو بسبب القيود في البنية التحتية، والقيود المتعلقة بالبيانات، ومشاكل الشفافية الخوارزمية، ومخاوف المعلمين بشأن العدالة والاستقلالية. ركزت معظم الدراسات على الأداء الأكاديمي (32.9٪) والتطبيقات المتعلقة بالصحة (22.4٪)، بينما لا تزال الجوانب الأخلاقية والشمولية غير مستكشفة بشكل كاف. **الاستنتاجات:** تمتلك أدوات التقييم المدعومة الذكاء الاصطناعي إمكانات كبيرة لتحويل ممارسات التقييم من خلال تعزيز العدالة والدقة ودعم التعلم الشخصي. ومع ذلك، يتطلب التنفيذ الفعال معالجة جاهزية المؤسسات، وتحسين كفاءة المعلمين، وضمان عمليات اتخاذ القرار الشفافة للحفاظ على الثقة والمساءلة. **التوصيات:** يجب أن تركز الأبحاث المستقبلية على التقييمات طويلة الأمد في الفصول الدراسية، وتطوير نماذج الذكاء الاصطناعي المتجاوبة ثقافيا، والتحقيق الأوسع في المخاطر الأخلاقية مثل التحيز وخصوصية البيانات. يشجع المؤسسات التعليمية وصناع السياسات على دمج أطر التقييم الموجهة الذكاء الاصطناعي مع الحكم التربوي الذي يقوده الإنسان، لضمان استخدام عادل ومسؤول لتقنيات الذكاء الاصطناعي في التعليم.

الكلمات المفتاحية: الذكاء الاصطناعي، التقييم بمساعدة الذكاء الاصطناعي، التكنولوجيا التعليمية، التغذية الراجعة الشخصية، التحيز في التقييم، الابتكار التعليمي، الذكاء الاصطناعي في التعليم.

1 Universitas Bengkulu, Indonesia
* Corresponding author email: eko_risdianto@unib.ac.id
2 INTI International University, Malaysia
3 University of Malaya, Kuala Lumpur, Malaysia
4 Department of Economics, The New College (University of Madras), Chennai, Tamil Nadu, India
5 Federal University of Kashere Gombe State Nigeria
6 CV Media Inti Teknologi, Indonesia

1 جامعة بنغكولو، إندونيسيا
* الباحث المراسل: eko_risdianto@unib.ac.id
2 جامعة إنتي الدولية، ماليزيا
3 جامعة مالايا، كوالالمبور، ماليزيا
4 قسم الاقتصاد، كلية نيو كوليدج (جامعة مدراس)، تشيناي، تاميل نادو، الهند
5 الجامعة الفدرالية في كاشيري، ولاية غومبي، نيجيريا
6 شركة ميديا إنتي تكنولوجي، إندونيسيا

Introduction

The development of artificial intelligence (AI) in the world of education has brought fundamental changes in the way learning and teaching take place (AMITABH KUMAR, 2024). AI has the potential to improve educational efficiency, personalization, and accessibility (Wang *et al*., 2024). Various studies have shown that the application of AI in education not only facilitates learning but also introduces challenges in achieving its maximum benefits (Vieriu & Petrea, 2025; Amoo, 2025). Additionally, AI plays a role in enhancing the efficiency of educational administration (Hutami, 2024; Murti *et al*., 2025). In this context, it is important for educators to adapt to the use of new technologies as well as understand their changing role in the AI-dominated educational ecosystem (Olanike Abiola Ajuwon *et al*., 2024). The application of AI-based education has the potential to drive the holistic development of students' abilities, including critical and creative skills (Su *et al*., 2023).

Recent empirical evidence also suggests that the use of generative AI in higher education is shaped by complex social and institutional factors, rather than being driven solely by technological capabilities. For instance, Hamamra, Mayaleh, and Khlaif (2024) demonstrated that students and educators negotiate the use of generative AI tools based on concerns regarding academic integrity, trust in automated outputs, and perceived cultural alignment with academic norms. These insights indicate that the integration of AI-assisted assessment requires careful consideration of contextual and pedagogical factors, not only technical implementation.

Although several prior systematic literature reviews have examined the use of AI in education, such as Wang *et al*. (2024), who focused on the general integration of AI in learning environments, and Garzón *et al*. (2025), who highlighted both benefits and challenges of AI adoption in classrooms, these reviews did not specifically analyze AI-assisted assessment tools in depth. This study extends the existing literature by combining bibliometric mapping with a systematic literature review approach, allowing for both a macro-level understanding of research trends and a micro-level evaluation of the effectiveness and challenges of AI-based assessment. As a result, this research offers a more comprehensive perspective on how AI is reshaping contemporary assessment practices in education.

Artificial Intelligence (AI) has demonstrated significant potential in enhancing efficiency, objectivity, and personalization in student assessment across various levels of education (Vieriu & Petrea, 2025) (González-Calatayud *et al*., 2021). One way AI supports assessments is by providing automated and objective systems, which reduce human bias and improve the accuracy of assessment results (Seneviratne & Manathunga, 2025). Automated scoring systems based on AI offer more consistent and accurate evaluations (Flament *et al*., 2021). As an assessment tool, AI can save educators time and enable them to focus more attention on developing more effective pedagogy (Kaldaras *et al*., 2024). Automated assessments can also help educators improve transparency and fairness in assessments, which are often challenges in traditional assessment systems (Reinhold *et al*., 2025)(Ragolane & Patel, 2024).

Various studies have shown that AI not only improves efficiency in the learning process but also personalizes the student learning experience, improving interaction between students and teachers (Gligorea *et al*., 2023) (Rincon-Flores *et al*., 2024). One of the latest innovations in the application of AI is the use of adaptive learning systems, which aim to

improve the effectiveness of teaching and the student experience (Guerrero-Roldán *et al*., 2021). The system adjusts teaching methods based on the needs and progress of each student, allowing for a more personalized and relevant approach to education (Cha *et al*., 2024).

Traditional education grading systems face several significant challenges, including inaccuracies, assessment bias, a lack of personalization, and an inability to accommodate the diverse needs of students (Cain *et al*., 2022) (Li *et al*., 2025). These challenges contribute to the dissatisfaction that both educators and learners often feel. First, inaccuracies in assessments are often caused by inconsistent evaluation methods, such as paper-based exams that can be influenced by a variety of factors, including student anxiety and coincidence. While modern technology can help teachers gain access to more accurate data, many assessments are still conducted manually and are prone to subjective errors (Swiecki *et al*., 2022; Dwikat, 2025). Artificial Intelligence (AI) has demonstrated significant potential (Mandava, 2025) in enhancing efficiency, objectivity, and personalization in student assessment across various levels of education.

Artificial intelligence (AI) has increasingly influenced educational practices, particularly in relation to teaching, learning, and assessment. Its use offers the potential to increase efficiency, reduce human bias, and provide tailored feedback based on individual learner needs. However, effective adoption requires educators to adapt to evolving technological environments while redefining their roles in AI-integrated ecosystems (Vieriu & Petrea, 2025; Hutami, 2024). Despite its benefits, the implementation of AI in education also introduces challenges related to data quality, transparency of algorithms, and institutional readiness for change.

The need to replace or complement traditional grading systems with more advanced technology is becoming increasingly urgent, given the challenges faced by current grading systems, including inaccuracies, bias, and a lack of personalization (Kooli & Yusuf, 2025; Al-Aghbari *et al*., 2025). New technologies, especially data-driven ones, can address these challenges by providing a more equitable, accurate, and responsive assessment of diverse student needs. One of the main challenges faced by traditional grading systems is the presence of inaccuracies and biases within them. Assessments that rely on manual methods are often prone to subjective errors, which can affect student assessment outcomes (Crogman *et al*., 2023) (Wei Wu, 2024).

A recent review published in Indonesia indicated that although AI technologies are increasingly being applied in project-based learning, the convergence with assessment practices remains limited (Zadeh *et al*., 2025). Artificial Intelligence (AI) offers numerous benefits in educational assessment, including assessment automation, personalized feedback, and more accurate data-driven analysis. The implementation of these innovations not only aims to improve efficiency but also to provide better and fairer assessment outcomes for all students (Saputra *et al*., 2024) (Garzón *et al*., 2025). However, while this technology promises a wide range of benefits, its implementation also faces several challenges, including data limitations and resistance to change (Khlaif *et al*., 2025; Jawabreh & Itmazi, 2025). Therefore, this study aims to identify current trends in the use of AI in educational assessment, identify the key authors and institutions involved, and explore the challenges faced in the application of AI-based assessment tools.

Methodology

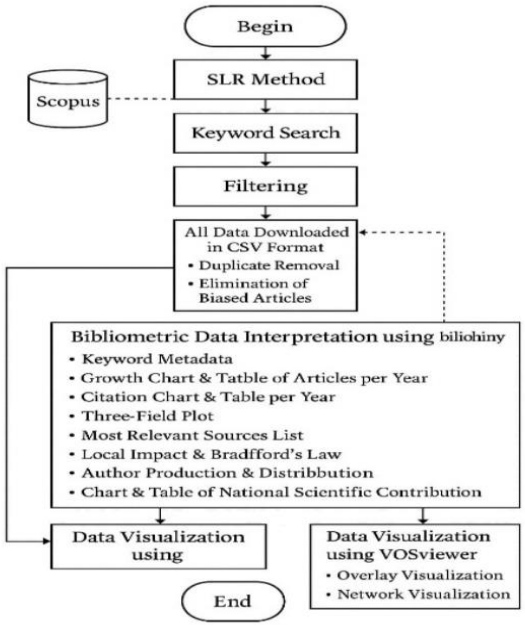

Figure (1): Research method flow diagram.

This study uses two main approaches, namely bibliometric analysis to explore publication trends and systematic literature review (SLR) to evaluate the effectiveness and challenges of artificial intelligence (AI)-based assessment tools in education (Suryanti *et al*., 2024). The methodological process is described as follows:

Bibliometric Approach

A bibliometric approach is used to analyze and map publication trends, key contributors, and relationships between authors and institutions (Sarı & Aypay, 2024) (Passas, 2024) in this topic. The measures of bibliometric analysis include database selection and data collection from some of the leading databases indexed by Scopus, due to the extensive coverage of the academic literature related to educational technology and AI.

Prism Method

The prism method is employed to conduct the data screening process (Pessin *et al*., 2023), the following are the steps used.

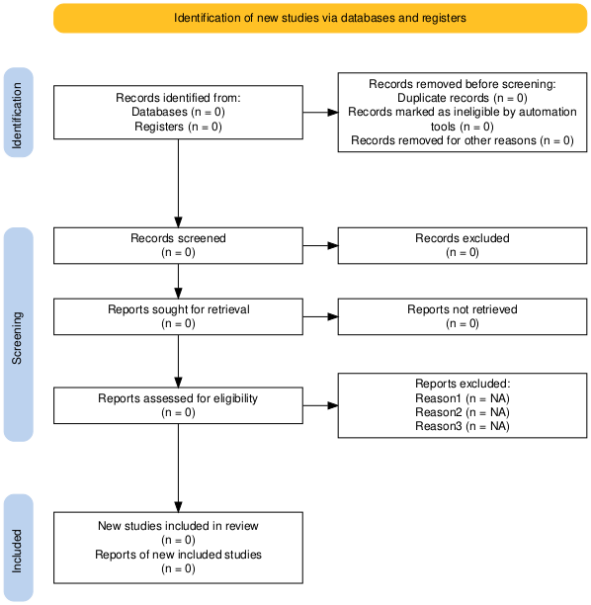

Figure (2): Prism method.

To ensure the reliability and validity of the review process, several bias-control and data validation procedures were implemented. First, duplicate records were identified and removed using OpenRefine, which also standardized metadata fields to avoid keyword-inflation bias. Second, the inclusion and exclusion criteria were applied consistently to maintain thematic coherence and to ensure that only studies directly related to AI-assisted assessment were considered. Third, only peer-reviewed journal publications indexed in Scopus were included, enhancing the academic rigor and credibility of the dataset. Finally, the combination of bibliometric analysis and systematic literature review allowed triangulation between quantitative patterns and qualitative thematic interpretation, thereby minimizing subjective bias.

Search Keywords

The search process uses the keyword "AI-assisted assessment tools" with inclusion criteria in the form of articles published in peer-reviewed journals, published since 2015, and accessible in a language understood, while exclusion criteria include articles that focus only on theory with no practical application, are irrelevant to educational assessment, or are not fully accessible. To meet this, in this study, 3

keyword combinations will be used to get articles that meet the criteria.

Keyword 1: Combinations for AI-Powered Grading Systems: ("Artificial Intelligence" OR "AI" OR "Machine Learning") AND ("Automated grading" OR "AI-assisted grading" OR "Intelligent grading system") AND ("Education" OR "E-learning" OR "Online learning") 21 dates. 2015-2025.

Keyword 2: Combination for Chatbot and AI-Based Feedback in Assessment: ("Artificial Intelligence" OR "AI" OR "Machine Learning") AND ("Chatbot" OR "Conversational agent" OR "AI tutor") AND ("Assessment" OR "Feedback system" OR "Automated feedback") AND ("Education" OR "Online learning") 106 of 2015-2025 data

Keyword 3: Combination for AI in Learning Analytics and Educational Assessment: ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning") AND ("Learning analytics" OR "Educational analytics" OR "Data-driven assessment") AND ("Student performance" OR "Student evaluation") 91 data from 2015-2025. From that total in 21+106+91=218

Reduce duplicate data

With the use of data from the three files of the above keyword search results, which are exported, then decompressed to a zip, and subsequently put into **Open Refine.** This application is used to filter data, particularly to determine whether data is obtained from search results. Additionally, this application can also be used to eliminate keyword bias contained in the data obtained. Initially, the data consisted of 218 articles, but after being entered into Open Refine, the number was reduced to 216 articles.

Metadata	Description	Missing Counts	Missing %	Status
AB	Abstract	0	0.00	Excellent
AU	Author	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
C1	Affiliation	1	0.46	Good
CR	Cited References	1	0.46	Good
DE	Keywords	14	6.48	Good
RP	Corresponding Author	53	24.54	Poor
ID	Keywords Plus	96	44.44	Poor
LA	Language	216	100.00	Completely missing
WC	Science Categories	216	100.00	Completely missing

Figure (3): The metadata on DE/Keyword is good.

After eliminating duplicate data and ensuring that the data is bias-free, the next step is to interpret the bibliometric data using Bibliosiny. Table 1 shows the main informational data that covers the entire article.

Table (1): Key Information on the data sheet.

Description	Results
Main Information About Data	
Timespan	2015:2025
Sources (Journals, Books, etc.)	140
Documents	216
Annual Growth Rate %	31.95
Document Average Age	2.06
Average citations per doc	33.82
References	10087
Document Contents	
Keywords Plus (EN)	1122
Author's Keywords (DE)	701
Authors	
Authors	1255
Authors of single-authored docs	11
Authors Collaboration	
Single-authored docs	12
Co-Authors per Doc	6.07
International co-authorships %	33.8
Document Types	
Article	175
Conference Paper	19
Reviews	22

As shown by the bibliometric data in Table 1, research on this subject has been collected from 2015 to 2025, comprising 216 documents from 140 sources, including journals, books, and other publications. The study is experiencing rapid growth, with an annual growth rate of 31.95%, indicating that the number of publications continues to increase annually.

Interestingly, the average age of the document is still relatively young, at approximately 2 years old, which suggests that this topic remains new and is still widely discussed. Each document was cited 33.82 times on average, indicating that its influence in academic circles was substantial. More than 10,000 references were used, demonstrating that this research was conducted with considerable accuracy and reliability. In terms of content, the authors identified 701 direct keywords and 1,122 additional keywords (Keywords Plus), highlighting the diversity of terms used in this study.

With 1,255 people involved, there were only 11 documents written by a single person, and the rest were the result of collaboration. This illustrates the importance of collaboration in advancing knowledge in this field.

This data shows that most research in this area is done in groups. Each document typically has an average of six writers, indicating that most of the writing is the result of teamwork. In addition, it is interesting that about 33,8% of the existing documents involve international collaborations, indicating that this topic is attracting the attention of researchers from all over the world. The majority of publications consist of journal articles (175 documents), followed by 19 conference papers and 22 review articles. As the dominance of these journal articles indicates, researchers more often use scientific journals as the primary medium for disseminating their research results than conferences or other types of media. Overall, we can see that this discipline is growing rapidly, thanks to the numerous collaborations among researchers, and it is gaining increasing attention in the academic world.

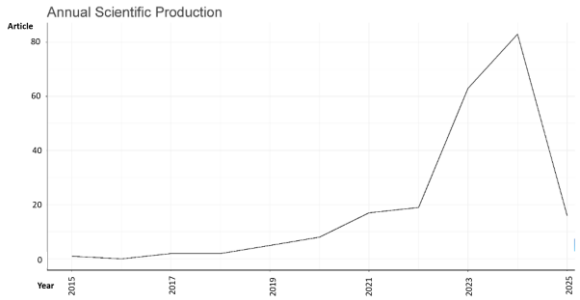

Figure (4): Annual growth chart.

This can be demonstrated by the data distribution in the table below.

Table (2): Annual Growth.

Year	Articles
2015	1
2016	0
2017	2
2018	2
2019	5
2020	8
2021	17
2022	19
2023	63
2024	83
2025	16

The number of articles published is expected to increase from 2015 to 2025, as indicated in Figure 4 and Table 2. Overall, the publication of articles has increased tremendously. In 2015, only one article was published, while in 2016, no article was published at all. In 2017 and 2018, the number of articles remained low, with two articles published each year. However, there has been a progressive increase starting in 2019. The number of articles reached five in 2019 and continues to grow to eight in 2020. With 17 and 19 articles, this significant surge continued in 2021 and 2022. With 63 and 83 articles, 2023 and 2024 set record highs, indicating a great interest in scientific production. Although the number of articles projected for 2025 is decreasing compared to the previous year, the figures show that they are still being published. The accompanying images provide a clearer picture of a rapidly increasing trend, particularly over the past few years. Increased interest or ability in relevant research areas is also indicated.

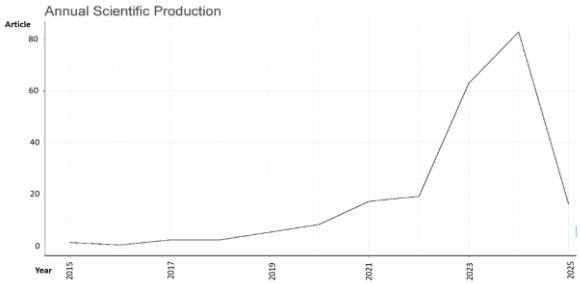

Figure (5): Annual Citation Chart.

It can be explained in the form of a table distribution as below.

Table (3): Annual Citation.

Year	Mean TC per Art	N	Mean T Cper Year	Citable Years
2015	166.00	1	15.09	11
2017	95.50	2	10.61	9
2018	27.00	2	3.38	8
2019	84.40	5	12.06	7
2020	38.00	8	6.33	6
2021	33.24	17	6.65	5
2022	25.05	19	6.26	4
2023	72.97	63	24.32	3
2024	6.31	83	3.15	2
2025	0.44	16	0.44	1

The graphs and Tables shown in Figure 5 and Table 3 display the average number of citations per year for articles published between 2015 and 2025. The graph illustrates the variation in the average citations, with 2015 reaching a peak, averaging 166 citations per article. However, until 2018, the average citation experienced a significant decline, dropping to 27.00, before increasing slightly again in 2019 to an average of 84.40. The years 2020 and 2021 showed relatively stable average citations, at 38.00 and 33.24, respectively. However, in 2022, there was another decline to 25.05, before jumping significantly to 72.97 in 2023. Unfortunately, the projections for 2024 indicate another drastic decrease in the average citation to 6.31, and it is expected to drop again in 2025 to 0.44. This data shows variations in how often the article is referenced by other researchers, illustrating the fluctuating impact of publications from year to year. Nonetheless, the annual citation totals show that the relevance and influence of the article are not always consistent; Factors such as the number of publications and the themes covered can affect this.

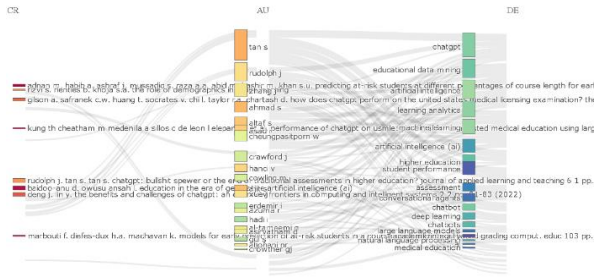

Figure (6): Three-Field Plot diagram.

Figure 6 shows a visualization of the relationship between the three main research elements of the cited author (CR), the author of the article (AU), and the main topic or keyword discussed (DE). This three-region visualization, also known as a three-region plot, helps illustrate the relationships between writers and ideas. Simply put, this image illustrates the authors who are often referenced (on the left side), the authors who write the articles (in the middle), and the topics discussed most frequently (on the right side). The lines connecting the columns indicate that the author is quoting other authors and discussing a specific topic. For example, if a line connects "Rudolph J" with "ChatGPT", it indicates that the author wrote an article discussing ChatGPT, possibly also citing the author in the left column. Overall, this image illustrates the interconnectedness of references, topics, and authors within the area of research being discussed. It is clear that the topic of AI in education, particularly ChatGPT, is a dominant one.

For the most relevant sources, they can be shown in the image below.

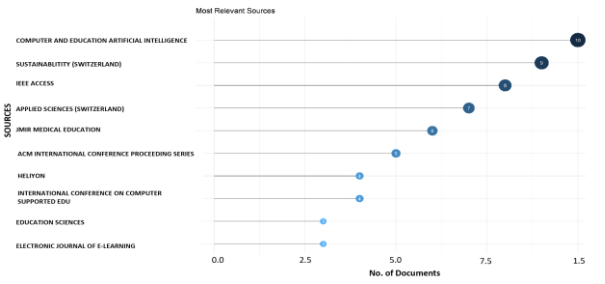

Figure (7): Most relevant Source.

Figure 7 shows the sources or journals that were most frequently used in the study; In general, this is a list of the most common "places" to publish articles related to the research subject. According to the figure, the journal "Computers and Education: Artificial Intelligence" is the most commonly cited, with 10 documents. It is followed by "Sustainability (Switzerland)" with nine articles, and "IEEE Access" with eight articles. Lower journals, such as the "Electronic Journal of E-Learning" and "Education Sciences", each contribute only one article. In addition, the size of the circle indicates the number of documents present; If the color is darker and larger, then more articles are coming from that source. Overall, this image shows all the most relevant or most frequently used journals as publication venues for research in this field. In other words, it is a list of researchers' "favorite journals" on their research topics. Overall, this image shows all the most relevant or most frequently used journals as publication venues for research in this field. In other words, it is a list of researchers' "favorite journals" on their research topics.

Next is the Sources' Local Impact data. This data is taken from the top 10.

Table (4): Sources' Local Impact.

Source	h_index	g_index	m_index	TC	NP	PY_start
Applied Sciences (Switzerland)	7	7	1.167	333	7	2020
Sustainability (Switzerland)	7	9	1.000	343	9	2019
Ieee Access	6	8	0.750	215	8	2018
Computers And Education: Artificial Intelligence	5	9	1.667	98	10	2023
Heliyon	4	4	1.333	74	4	2023
Jmir Medical Education	4	6	1.333	1069	6	2023
Acm International Conference Proceeding Series	3	5	0.750	40	5	2022
Education Sciences	3	3	1.000	785	3	2023
Journal Of Applied Learning And Teaching	3	3	1.000	1187	3	2023
Medicine (United States)	3	3	1.500	12	3	2024

The table shows more detailed data on the most important sources of publications or journals in this field of research. Table 4 presents some of the bibliometric metrics for each journal. For example, the h-index shows the number of articles that are cited at least h times. The G-index is comparable to the h-index but prioritizes more frequently cited articles. And the M-index shows the average development of the h-index per year since the journal began publishing articles in this field. Total citations (TC) show the total number of citations received from all journal documents. The number of publications refers to the total number of articles published. PY_start, the year in which journals began publishing articles on the topic.

Next is the Core Sources by Bradford's Law data. This data is taken from the top 10.

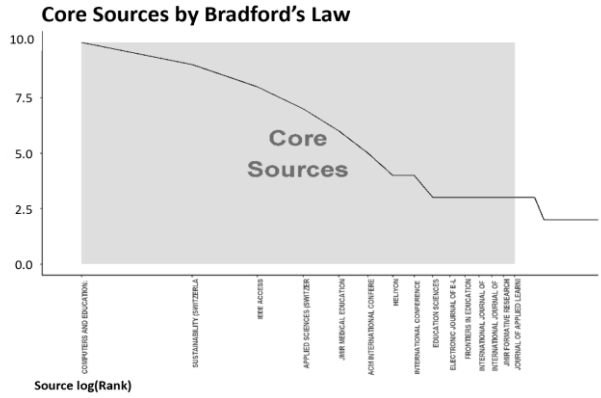

Figure (8): Core Sources by Bradford's Law.

Figure 8 shows the "primary sources" in the research area that are based on Bradford's Law. Using this graph, we can find the journals or sources that are most frequently used or most important for the research topic. The black area, known as the "Core Source," is where the most important and useful sources are gathered. Researchers in this field most often use these journals as references. For illustrations, see

journals such as Computer and Education, Artificial Intelligence, Sustainability (Switzerland), IEEE Access, and Applied Sciences (Switzerland), among others, on the left. Because they account for a sizable number of articles compared to other journals, these journals fall under the main or core group. When you look to the right, you'll see a significant decrease in the number of articles provided by each source. This indicates that, although more sources are used, their contribution to the total number of articles is decreasing. This is the essence of Bradford's Law: one journal contributes a lot of articles, and the rest contribute little. Essentially, this graph enables us to identify which journals are the most significant or influential in disseminating knowledge related to a specific topic. Journals that are in the "primary source" field are the best places to start looking for references if your focus is on AI literature in education.

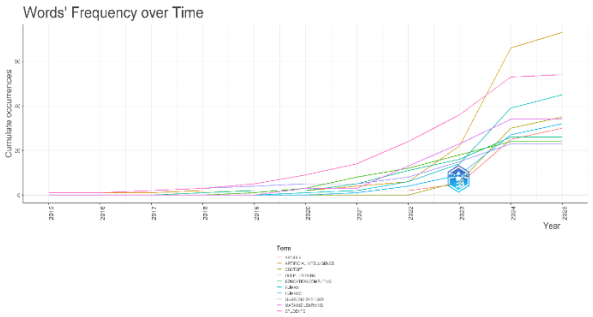

Figure (9): Cumulate Occurrences.

The use of keywords in academic publications increased from 2015 to 2025, as shown in Figure 9. Since 2023, there has been a significant increase in the terms "Artificial Intelligence" and "ChatGPT". This illustrates the growing interest in this topic within the realm of research as AI technology continues to advance. Additionally, terms such as "Machine Learning", "Deep Learning", and "Learning Systems" are constantly evolving. Meanwhile, more general words such as "student" and "human" continue to emerge, suggesting that the focus of this research is still on how technology affects humans and learning. Even the word "Article" is becoming more popular, indicating that many literature studies have been conducted. Overall, this graph shows that interest in research connecting smart technologies to education has increased substantially in recent years. This trend is likely to continue in the future.

Data authors Production over Time.

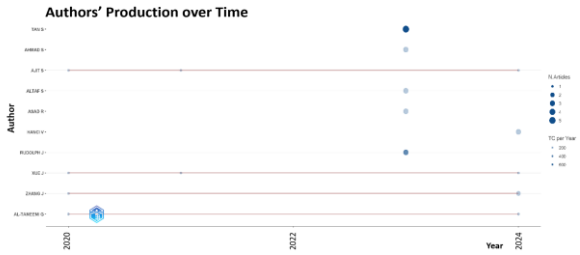

Figure (10): Authors' Production over Time.

The author's production from 2020 to 2024 is depicted in Figure 10. With five articles published in one year and a fairly high citation rate, Tan S. seems to be the most prolific writer. Other authors, such as Ahmad S., Ajit S., Altaf S., and Asad R., have also made consistent contributions, especially in 2023, which seems to be the most active year in terms of publications. Meanwhile, some writers, such as Rudolph J. and Hanci V., remain relevant despite having written little. This graph shows the number of articles published, while the color shows the intensity or number of citations per year. This allows us to see which authors are not only productive but also have quite significant works. According to this graph, scientific collaboration and productivity have increased in recent years, especially since 2023.

Countries

The following is the Author's Corresponding by country.

Figure (11): Author's Corresponding by country.

The most active authors of the publication's correspondents are shown in Figure 11. It is clear that the United States (US) has the largest number of documents, and most of the publications result from international cooperation (indicated in blue). Saudi Arabia and the UK are also active, but Saudi Arabian publications are more locally produced (Single Country Publications (SCPs) than British publications, which usually collaborate. Other countries, such as China, Canada, Pakistan, and Turkey, also actively utilize local and international publications. Cross-border collaboration is essential for scientific production, and this graph shows that research in this area is being conducted in many countries. In summary, this graph illustrates the growing contribution of this field of research worldwide and the strong trend toward international cooperation.

The following is the data of the top 10 Countries' Scientific Production.

Table (5): Top 10 Countries' Scientific Production.

Country	Freq
USA	397
UK	73
CHINA	70
AUSTRALIA	55
CANADA	42
SAUDI ARABIA	39
ITALY	37
PAKISTAN	34
TURKEY	32
INDIA	29

The list of countries with the most publications on the field studied is shown in table 5. The United States (USA) has the highest number of publications with 397, far surpassing any other country. The UK is second with 73 publications, followed by China with 70 publications. Countries such as Australia, Canada, and Saudi Arabia are also very active, with dozens of articles created. In addition, India, Pakistan, Turkey, and Italy are in the top ten. Overall, this graph shows that, despite the scientific contributions spread across different countries, the United States seems to have the largest number of publications.

Countries' Production over Time

The following is the data on Countries' Production over Time. The data can be seen in the graph below.

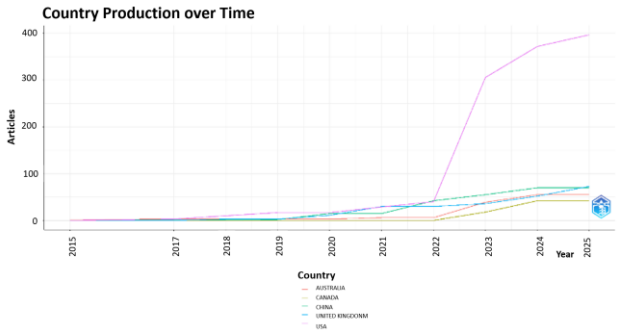

Figure (12): Charts of Countries' Production over Time.

Figure 12 shows the trend in the number of published articles from the five countries of the United States, the United Kingdom, China, Canada, and Australia during the period 2015 to 2025. It can be seen that the United States (USA) dominates the number of publications significantly, especially after 2022, when there was a huge surge that reached 300 articles and will continue to increase to 400 articles by 2025. The United Kingdom, Canada, Australia, and China also show an upward trend, but on a smaller scale. The sharp surge in the United States shows that in recent years, the country has become the center of attention or focus of major research in this field.

Most Cited Countries

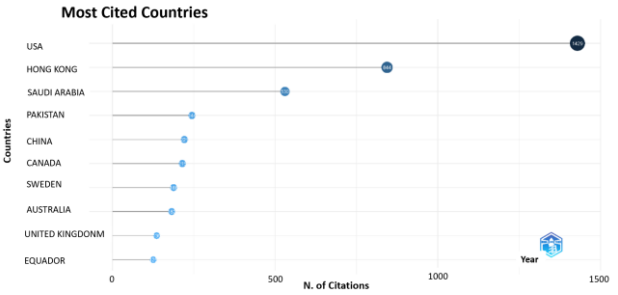

Figure (13): Most Cited Countries.

A list of countries whose articles are most frequently cited by other researchers is shown in Figure 13. The United States (US)

outperformed all other countries with the highest number of citations, 1,429 times, as this graph shows. Hong Kong ranked second with 844 citations, while Saudi Arabia followed in third place with 530 citations. Other countries, such as Pakistan, China, and Canada, are also on the list, but the number of citations is lower. This demonstrates, as other researchers widely acknowledge, that research contributions from the US and Hong Kong have a significant impact on the field.

Overlay visualization.

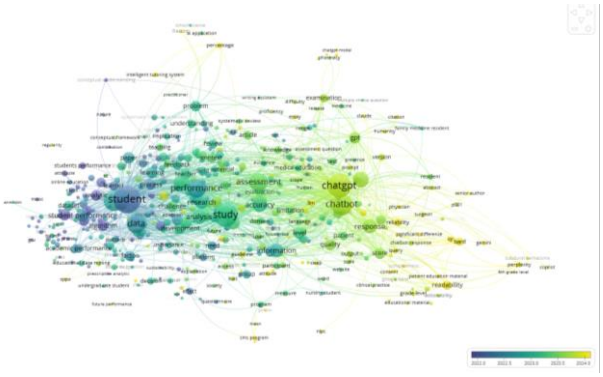

Figure (14): Contact map.

Figure 14 shows the relationship of keywords from various publications, with the year of the keyword's appearance indicated in color. On the left, words like "student", "performance", and "data" dominate, indicating long-discussed topics, such as "ChatGPT", "chatbot", and "response". On the right side, words like "response", "chatbot", and "student" appear brighter (yellow), indicating that they are related. Additionally, the network between words illustrates the relationship between one topic and another. For example, the network between words shows the relationship between "evaluation" and "thoroughness" or "ChatGPT" and "readability". Overall, this graph shows the shift in research focus from an emphasis on student performance to the use of technologies such as ChatGPT in the context of education or health.

Network Visualization

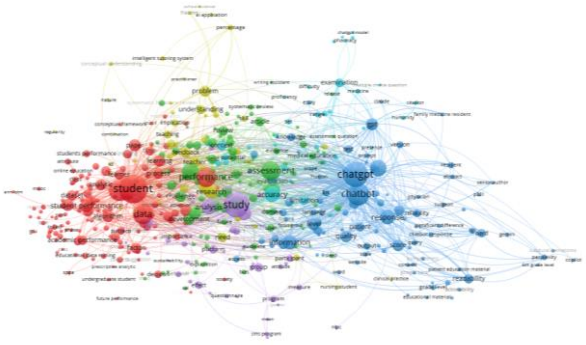

Figure (15): Network Visualization.

Figure 15 illustrates a visual map of keywords that frequently co-occur in various publications, where each color represents a group of interrelated topics. On the left side, the red group focuses on themes such as "student", "performance", and "data", which shows a lot of research related to education and student performance analysis. In the middle, the green group discusses things like "assessment", "accuracy", and "study", describing more general topics of evaluation and research. Meanwhile, on the right side, the blue group dominates with words like "ChatGPT", "chatbot", and "response", reflecting the latest technological trends in education and communication. This picture, as a whole, gives an idea of how research that initially focused on education is now starting to connect significantly with AI technology, such as ChatGPT.

Systematic Literature Review (SLR) Approach

SLR is used to systematically analyze selected articles, evaluate the effectiveness of the use of AI in educational assessments, and identify implementation challenges faced.

Of the 218 data obtained, 174 will be focused on the type of article.

Table (6): SLR Study Data.

No	Type	Number
1	Article	174
2	Conference Paper	19
3	Reviews	23
Total		**216**

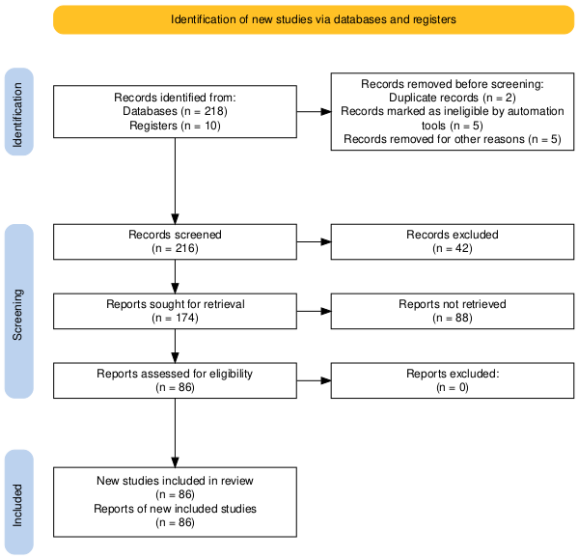

Figure (16): Prism Method Diagram.

Figure 16 illustrates the process of identifying new studies through databases and registries, which began by collecting a total of 218 records from various databases and 10 records from registries. Before filtering, two records were deleted due to duplication, five records were marked as ineligible by the automated tool, and five records were deleted for other reasons. After screening, a total of 216 records were examined, of which 42 records were ultimately excluded from further consideration. Of the 174 reports sought, 88 were unsuccessful. Of the remaining reports, 86 reports were assessed for eligibility, and no reports were excluded in this process. Finally, 86 new studies were successfully included in the review, creating a valid dataset for further analysis. This process demonstrates a systematic approach to collecting and processing research information.

The findings of this review affirm that AI-assisted assessment tools contribute to improved scoring consistency, efficiency, and personalized feedback. This aligns with previous evidence that automated scoring systems can reduce human subjectivity in evaluation processes (Seneviratne & Manathunga, 2025). However, the results also indicate that the adoption of AI-based assessment is not uniform across educational contexts. Institutions with limited digital infrastructure face implementation barriers, while educators may express concerns regarding loss of autonomy when automated systems influence grading decisions. These findings suggest that AI should be positioned not as a replacement for teacher expertise, but as a decision-support mechanism designed to enhance pedagogical judgment and assessment fairness.

Table (7): Key Classifications in the Elicit AI Methodology Table.

Methodology Categories	Number of Papers	Percentage (%)
Experimental Studies	12	14.1%
Observational & Quantitative Studies	15	17.6%
Systematic Reviews & Literature Reviews	10	11.8%
Qualitative Research	8	9.4%
Computational Studies & AI Model Analysis	9	10.6%
Case Studies & Descriptive Studies	11	12.9%
Protocols & Conceptual Frameworks	7	8.2%
Others/Unclear	13	15.3%
Total	**85**	**100%**

The research methodological classification of AI chatbots shows the dominance of observational-quantitative (17.6%) and experimental (14.1%) approaches, which together account for 32%, reflecting the main focus on empirical validation. Other methodologies that are quite prominent include case studies and descriptive (12.9%), systematic reviews (11.8%), and computational studies related to AI models (10.6%). Meanwhile, qualitative research accounts for only 9.4%, indicating that there is room for a deeper exploration of user perceptions and experiences. The "miscellaneous/unclear" category accounted for 15.3% of the total 85 publications, which may reflect a lack of explicit methodological reporting.

Table (8): Key Classifications in the Elicit AI Main Findings Table.

Main Findings Category	Number of Papers	Percentage (%)	Examples of Specific Findings
Academic Performance	28	32.9%	- GPT-4 excels in medical exams Chatbots improve vocabulary retention
Health Applications	19	22.4%	- 92% accuracy for rosacea information - Diagnostic support capabilities
Technical Limitations	15	17.6%	- Hallucinations in 23% of cases - Readability above high school level
User Acceptance	12	14.1%	- 78% of students satisfied with AI tutors - 62% of doctors feel comfortable
System Efficiency	7	8.2%	- Reduces grading workload by 40% - Automates 35% of administrative tasks
Ethical Implications	4	4.7%	- 31% risk of plagiarism - Gender bias in responses

The classification of key findings from 85 studies on AI chatbots shows a dominant focus on education and health, with the categories of academic performance (32.9%) and health applications (22.4%) combined accounting for more than half of the findings (55.3%). Other findings included technical limitations (17.6%), such as hallucinations and high readability, as well as user acceptance (14.1%) which remained positive despite technical issues. Aspects of system efficiency were identified in only 8.2% of studies, while ethical issues, such as plagiarism and gender bias, were highlighted least frequently (4.7%). Overall, the majority of studies (73%) emphasized technical effectiveness over exploring social and ethical impacts (18.8%), indicating a need for a balance in the direction of future research.

Table (9) Key Classifications in the Elicit AI Summary Table.

Category	Number of Papers	Percentage (%)	Examples of Key Findings
Education	31	36.5%	- GPT-4 outperforms 98% of medical students - Increases vocabulary retention by 40%
Health	22	25.9%	- 92% accuracy for chronic disease information - Reduces doctors' workload by 30%
Technical Evaluation	18	21.2%	- Hallucination rate of 15-25% - Readability two levels above recommendations
Business Implementation	7	8.2%	- Increases customer service efficiency by 25% - Reduces operational costs by 35%
Social-Ethical Aspects	5	5.9%	- 31% risk of academic plagiarism - Gender bias in 18% of responses
Others	2	2.4%	- Specialized applications for disabilities - Use in legal research

A summary of classifications from 85 studies on AI chatbots shows the dominance of studies in the fields of education (36.5%) and health (25.9%), which together account for more than half of the total publications. The main focus of the research seems to be on improving efficiency and accuracy, reflected in the categories of technical evaluation (21.2%) and business application (8.2%), so that a total of 83.6% of the studies are oriented towards system performance. Meanwhile, social and ethical aspects are still relatively underexplored (5.9%), and other categories only cover 2.4% of the research. Key findings include significant improvements in learning retention, accuracy of medical diagnoses, and operational efficiency, but also reveal potential risks such as plagiarism and response bias.

Table (10): Key Classifications in the Elicit AI Design Study Table.

Study Design	Number of Papers	Percentage (%)	Unique Characteristics	Examples of Application
Cross-Sectional	31	32%	Single time-point measurement	Educational chatbot user satisfaction survey
Experimental	17	18%	RCT/Quasi-experimental with control	Testing the effectiveness of AI tutors vs human tutors
Qualitative	14	15%	In-depth interviews, FGD	Lecturers' perceptions of ChatGPT
Longitudinal	11	12%	Repeated time measurement	The long-term impact of learning chatbots
Systematic review	10	10%	PRISMA, meta-analysis	Synthesis of evidence on health applications
Case/Control	8	8%	Specific comparative analysis	ChatGPT vs Bard in diagnosis
Description	5	5%	In-depth observation without intervention	Case study of LMS+AI implementation

The classification of study designs in AI chatbot research showed the dominance of cross-sectional designs (32%), which were chosen for their efficiency in collecting data at one time, such as user satisfaction surveys. The experimental design covered only 18% of the studies, with a small percentage using pure RCTs, reflecting the challenges of controlling variables in AI interventions. Qualitative approaches (15%) are starting to improve, especially to explore human-centered aspects, while longitudinal design (12%) is still relatively rare although it is important to assess long-term impacts. Systematic (10%), case/control studies (8%), and descriptive (5%) rounded out the diverse methodological landscape, with some studies incorporating more than one design.

Table (11): Key Classifications in the Future Research Table Elicit AI.

Category	Frequency	Level of Urgency	Representative Examples
Model Validation	27	High	Development of new algorithms
Longitudinal	21	High	Long-term impact (5+ years)
Ethics	17	Critical	AI regulatory framework
Education	14	Moderate	Integration of AI into the curriculum
Clinical	10	High	Medical decision support systems
Comparison	5	Moderate	Comparison of ChatGPT vs Bard
Inclusion	4	Emerging	Accessibility for people with disabilities
Economics	3	Emerging	Analysis of AI implementation costs

The "Future Research" classification in AI chatbot research shows that topics with a high level of urgency are most focused on model validation (27 studies) and longitudinal studies (21 studies), emphasizing the importance of developing new algorithms as well as evaluating long-term impacts. Ethical issues emerged as a critical category with 17 studies, reflecting the urgent need for an AI regulatory framework. Meanwhile, the educational (14 studies) and clinical (10 studies) aspects were considered quite important but did not the previous topic. Topics of comparative, inclusion, and economics are still classified as emerging trends, pointing to a more diverse future exploration direction, from accessibility to the cost efficiency of AI implementation.

Conclusions and Implications

This study concludes that AI-assisted assessment tools have significant potential to enhance educational evaluation by improving fairness, efficiency, and personalization. However, successful implementation requires addressing institutional readiness, transparent algorithmic decision-making, and educator capacity building. Policymakers should prioritize ethical regulatory frameworks, data protection, and equitable access to AI

technologies. Future research should focus on longitudinal classroom-based evaluations, culturally adaptive AI model development, and strategies for integrating AI feedback with human-led formative assessment practices. In addition, recent research has shown that educators' attitudes play a central role in determining the extent to which generative AI tools are integrated into assessment practices. Khlaif *et al*. (2024) reported that while many university instructors recognize AI's potential to support evaluation, concerns related to fairness, transparency, and preparedness continue to influence adoption decisions. These insights suggest that professional development and institutional policy frameworks are essential to support effective and responsible use of AI in assessment.

Disclosure Statement

The authors declare that this study was conducted independently and without any influence from external parties. All analyses, interpretations, and conclusions presented in this manuscript are solely based on the authors' academic work and do not reflect the views of any affiliated institutions or funding bodies. No external entity had control over the design, data selection, analysis, or reporting of this research.

- **Ethical Approval and Consent to Participate:** This study is based solely on bibliometric data and published scholarly articles. It does not involve human participants, personal data, or interventions requiring ethical clearance. Therefore, ethical approval and informed consent were not required. All data used in this study were obtained from publicly accessible scientific databases.
- **Availability of Data and Materials:** All data supporting the findings of this study are available from openly accessible sources, primarily the Scopus-indexed database. The bibliometric datasets generated and analyzed during this research (including metadata, keyword files, and refined datasets) can be provided by the corresponding author upon reasonable request.
- **Authors' Contribution:** Eko Risdianto contributed to the conceptualization, research design, data acquisition, bibliometric analysis, SLR integration, manuscript drafting, and overall supervision of the study. Rada Kristina supported data refinement, OpenRefine processing, visualization interpretation, and manuscript editing. Deshinta Arrova Dewi assisted in the systematic review process, classification table development, and analysis validation. Ahmad Zabidi Abdul Razak provided methodological consultation, critical review of theoretical frameworks, and enhancement of the discussion section. M. Abdul Jamal contributed to comparative analysis, statistical interpretation, and literature synthesis. Adamu Abubakar Muhammad reviewed the results, provided international contextualization, and assisted with manuscript proofreading. Hera Anis Kartika supported data processing, figure preparation, and administrative coordination throughout the research process.
- **Conflict of Interest:** The authors declare that there is no conflict of interest regarding the publication of this article. The authors have no financial, institutional, or personal relationships that could influence the work reported in this study.
- **Funding:** This research did not receive any specific grant from public, commercial, or not-for-profit funding agencies. The study was conducted using the authors' own resources and institutional support.
- **Acknowledgments:** The authors would like to express their gratitude to Universitas Bengkulu and the collaborating institutions

for providing academic support throughout this research. The authors also acknowledge the contributions of database providers and open-source tools such as Scopus, OpenRefine, and Biblioshiny, which enabled the bibliometric and systematic review processes. Special thanks are extended to colleagues who provided constructive feedback during the manuscript development.

Open Access

References

- Al-Aghbari, N., Alkharusi, H., & Al Fazari, M. (2025). AI-Based Formative Assessment and its Impact on the Academic Achievement and Perceptions of the Classroom Assessment Environment. *An-Najah University Journal for Research - B (Humanities), 39*(11), 893–906. https://doi.org/10.35552/0247.39.11.2484
- AMITABH KUMAR. (2024). Artificial Intelligence in Education: Revolutionizing Teaching and Learning. Journal of Asian Primary Education (JoAPE), 1(1), 64–68. https://doi.org/10.59966/joape.v1i1.1207
- Amoo, T. A. (2025). Ethical Consideration in Implementing AI-based Tutoring Systems as Educational Technology Tool in Education: Balancing Efficiency with Privacy and Equity in the Teaching of Students. FINGER : Jurnal Ilmiah Teknologi Pendidikan, 4(1), 34–41. https://doi.org/10.58723/finger.v4i1.365
- Cain, J., Medina, M., Romanelli, F., & Persky, A. (2022). Deficiencies of Traditional Grading Systems and Recommendations for the Future. American Journal of Pharmaceutical Education, 86(7), 908–915. https://doi.org/10.5688/ajpe8850
- Cha, S., Loeser, M., & Seo, K. (2024). The Impact of AI-Based Course-Recommender System on Students' Course-Selection Decision-Making Process. Applied Sciences (Switzerland), 14(9), 3672. https://doi.org/10.3390/app14093672
- Crogman, H. T., Eshun, K. O., Jackson, M., TrebeauCrogman, M. A., Joseph, E., Warner, L. C., & Erenso, D. B. (2023). Ungrading: The Case for Abandoning Institutionalized Assessment Protocols and Improving Pedagogical Strategies. Education Sciences, 13(11). https://doi.org/10.3390/educsci13111091
- Dwikat, H. (2025). The Impact of Artificial Intelligence on the Economy in the Media Industry: Perspectives of Palestinian Media Workers. *An-Najah University Journal for Research - B (Humanities), 39*(8), 611–626. https://doi.org/10.35552/0247.39.8.2453
- Flament, F., Lee, Y. W., Lee, D. H., Passeron, T., Zhang, Y., Jiang, R., Prunel, A., Dwivedi, S., Kroely, C., Park, Y. J., Chuberre, B., & Aarabi, P. (2021). The continuous development of a complete and objective automatic grading system of facial signs from selfie pictures: Asian validation

study and application to women of three ethnic origins, differently aged. Skin Research and Technology, 27(2), 183–190. https://doi.org/10.1111/srt.12922

– Garzón, J., Patiño, E., & Marulanda, C. (2025). Systematic Review of Artificial Intelligence in Education : Trends , Benefits , and Challenges. Multimodal Technologies and Interaction, 9(84), 1–19. https://doi.org/10.3390/mti9080084
– Gligorea, I., Cioca, M., Oancea, R., Gorski, A. T., Gorski, H., & Tudorache, P. (2023). Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review. Education Sciences, 13(12). https://doi.org/10.3390/educsci13121216
– González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. Applied Sciences (Switzerland), 11(12). https://doi.org/10.3390/app11125467
– Guerrero-Roldán, A. E., Rodríguez-González, M. E., Bañeres, D., Elasri-Ejjaberi, A., & Cortadas, P. (2021). Experiences in the use of an adaptive intelligent system to enhance online learners' performance: a case study in Economics and Business courses. International Journal of Educational Technology in Higher Education, 18(1). https://doi.org/10.1186/s41239-021-00271-0
– Hamamra, B., Mayaleh, A., & Khlaif, Z. N. (2024). Between tech and text: the use of generative AI in Palestinian universities - a ChatGPT case study. Cogent Education, 11(1), 1–14. https://doi.org/10.1080/2331186X.2024.2380622
– Hutami, S. (2024). Utilizing Technology and Artificial Intelligence in Educational Administration to Enhance School Performance at Junior High School. PPSDP International Journal of Education, 3(2), 197–212. https://doi.org/10.59175/pijed.v3i2.302
– Jawabreh, M., & Itmazi, J. (2025). Efficiency of Artificial Intelligence in Answering Questions of Palestinian General Secondary Education Exam (Tawjihi) in Arabic Curricula: Case of (ChatGPT) and (Gemini). *An-Najah University Journal for Research - B (Humanities), 39*(8), 551–570. https://doi.org/10.35552/0247.39.8.2442
– Kaldaras, L., Haudek, K., & Krajcik, J. (2024). Employing automatic analysis tools aligned to learning progressions to assess knowledge application and support learning in STEM. International Journal of STEM Education, 11(1). https://doi.org/10.1186/s40594-024-00516-0
– Khlaif, Z. N., Alkouk, W. A., Salama, N., & Abu Eideh, B. (2025). Redesigning Assessments for AI-Enhanced Learning: A Framework for Educators in the Generative AI Era. Education Sciences, 15(2). https://doi.org/10.3390/educsci15020174
– Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A. A., Ayyoub, A., Hattab, M. K., & Shadid, F. (2024). University Teachers' Views on the Adoption and Integration of Generative AI Tools for Student Assessment in Higher Education. In Education Sciences (Vol. 14, Nomor 10). https://doi.org/10.3390/educsci14101090
– Kooli, C., & Yusuf, N. (2025). Transforming Educational Assessment: Insights Into the Use of ChatGPT and Large Language Models in Grading. International Journal of Human-Computer Interaction, 41(5), 3388–3399. https://doi.org/10.1080/10447318.2024.2338330

– Li, T., Haudek, K., & Krajcik, J. (2025). Utilizing Deep Learning AI to Analyze Scientific Models: Overcoming Challenges. Journal of Science Education and Technology, 866–887. https://doi.org/10.1007/s10956-025-10217-0
– Mandava, R. (2025). Analyzing Bias in Large Language Models: A Quantitative Study Using Sentiment and Demographic Metrics. International Journal of Advances in Artificial Intelligence and Machine Learning. International Journal of Advances in Artificial Intelligence and Machine Learning, 2(2), 67–75. https://doi.org/10.58723/ijaaiml.v2i2.403
– Murti, Y. R., Ramadhani, D. P., & Irawan, H. (2025). Utilizing Retrieval Augmented Generation (RAG) -Based Chatbots as an Innovative Learning Tool in Higher Education : A Case Study on the Use of Digital Learning Resources. Indonesian Journal of Elearning and Multimedia, 4(3), 281–298. https://doi.org/10.58723/ijoem.v4i3.510
– Olanike Abiola Ajuwon, Enitan Shukurat Animashaun, & Njideka Rita Chiekezie. (2024). Integrating AI and technology in educational administration: Improving efficiency and educational quality. Open Access Research Journal of Science and Technology, 11(2), 116–127. https://doi.org/10.53022/oarjst.2024.11.2.0102
– Passas, I. (2024). Bibliometric Analysis: The Main Steps. Encyclopedia, 4(2), 1014–1025. https://doi.org/10.3390/encyclopedia4020065
– Pessin, V. Z., Santos, C. A. S., Yamane, L. H., Siman, R. R., Baldam, R. de L., & Júnior, V. L. (2023). A method of Mapping Process for scientific production using the Smart Bibliometrics. MethodsX, 11(September), 102367. https://doi.org/10.1016/j.mex.2023.102367
– Ragolane, M., & Patel, S. (2024). Transforming Educ-AI-tion in South Africa: Can AI-Driven Grading Transform the Future of Higher Education? Journal of Education and Teaching Methods, 3(1), 26–51. https://doi.org/10.58425/jetm.v3i1.267
– Reinhold, L., Händel, M., & Naujoks-Schober, N. (2025). AI-teacher agreement in evaluating learning diaries. Frontiers in Education, 10. https://doi.org/10.3389/feduc.2025.1601789
– Rincon-Flores, E. G., Castano, L., Guerrero Solis, S. L., Olmos Lopez, O., Rodríguez Hernández, C. F., Castillo Lara, L. A., & Aldape Valdés, L. P. (2024). Improving the learning-teaching process through adaptive learning strategy. Smart Learning Environments, 11(1). https://doi.org/10.1186/s40561-024-00314-9
– Saputra, I., Kurniawan, A., Yanita, M., Yeni Putri, E., & Mahniza, M. (2024). The Evolution of Educational Assessment: How Artificial Intelligence is Shaping the Trends and Future of Learning Evaluation. The Indonesian Journal of Computer Science, 13(6), 9056–9074. https://doi.org/10.33022/ijcs.v13i6.4465
– Sarı, T., & Aypay, A. (2024). A Bibliometric Study of Issues in Educational Policy. Education Sciences, 14(6), 1–23. https://doi.org/10.3390/educsci14060568
– Seneviratne, H. M. T. W., & Manathunga, S. S. (2025). Artificial intelligence assisted automated short answer question scoring tool shows high correlation with human examiner markings. BMC Medical Education, 25(1). https://doi.org/10.1186/s12909-025-07718-

- Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial Intelligence (AI) Literacy in Early Childhood Education: The Challenges and Opportunities. Computers and Education: Artificial Intelligence, 4(January 2023), 100124. https://doi.org/10.1016/j.caeai.2023.100124
- Suryanti, R., Jahidin, J., & Fadlil, M. (2024). Artificial Intelligence in Education: Bibliometric and Systematic Literature Review from 2019 – 2024. International Education Trend Issues, 2(2), 231–255. https://doi.org/10.56442/ieti.v2i2.647
- Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. Computers and Education: Artificial Intelligence, 3(May). https://doi.org/10.1016/j.caeai.2022.100075
- 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, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. Expert Systems with Applications, 252(PA), 124167. https://doi.org/10.1016/j.eswa.2024.124167
- Wei Wu. (2024). Development and Application of Big Data-Driven Intelligent Evaluation System for Physical and Aesthetic Education. Journal of Electrical Systems, 20(3s), 2447–2458. https://doi.org/10.52783/jes.3145
- Zadeh, M., Isaee, H., & Barjesteh, H. (2025). Artificial Intelligence in Project-Based Learning: A Systematic Review of Its Role in English Language Acquisition and Pedagogical Innovation. Indonesian Journal of Pedagogy and Teacher Education, 3(3), 81–91. https://doi.org/10.58723/ijopate.v3i3.502