Mapping the Research Landscape of AI-Driven Formative Assessment for Argumentation Skill Development in Higher Education: A Bibliometric Study

Ahmad Fauzi Hendratmoko [1], Madlazim Madlazim [2], Wahono Widodo [1], Elok Sudibyo [1], Mohammad Budiyanto [1], Sapti Puspitarini [1], Mohd Istajib Bin Mokhtar [3]

[1] *Department of Science Education, Universitas Negeri Surabaya, Surabaya, Indonesia*
[2] *Department of Physics, Universitas Negeri Surabaya, Surabaya, Indonesia*
[3] *Department of Science and Technology Studies, Universiti Malaya, Kuala Lumpur, Malaysia*

Abstract **– The integration of AI into formative assessments to enhance students' argumentation skills in higher education has gained growing attention over the past 30 years. However, research in this area remains fragmented. This study uses bibliometric analysis of 109 Scopus-indexed publications (1992–2025) with VOSviewer and Biblioshiny to map its evolution, key contributors, major themes, and future directions. Findings reveal a 6.5% annual growth rate, indicating sustained academic interest. Major contributions come from the USA, Germany, Switzerland, and the UK. Thematic clusters include: (1) AI and argumentation in education, (2) Instructional systems and e-learning, (3) Automated feedback and LLM-based tools like ChatGPT, and (4) NLP methods. Despite progress, significant gaps remain, such as limited attention to process-based assessments, underrepresentation of multilingual and collaborative settings, and concerns over the fairness of automated feedback. The study calls for interdisciplinary research, culturally responsive design, and longitudinal studies to fully leverage AI in advancing 21st-century argumentation skills.**

DOI: 10.18421/TEM153-59
https://doi.org/10.18421/TEM153-59

Corresponding author: Ahmad Fauzi Hendratmoko,
Department of Science Education, Universitas Negeri Surabaya, Surabaya, Indonesia
Email: ahmadhendratmoko@unesa.ac.id

Received: 20 September 2025.
Revised: 23 January 2026.
Accepted: 12 February 2026.
Published: 27 August 2026.

Keywords **– Argumentation skills, AI, formative assessment, higher education, bibliometric.**

1. Introduction

Artificial Intelligence (AI) is now used a lot more in higher education than it was in the past. Formative assessment is one way that this is put into practice. It helps learning processes that are based on feedback [1], [2], [3]. AI-based formative assessment allows for the delivery of automated, high-quality, and individualized feedback [4], [5], which may enhance student learning results.

One of the main goals of higher education is to teach students how to argue. This skill is recognized as an essential element of critical thinking, a competency for the 21st century [6], [7]. The development of argumentation skills is often impeded by lecturers' insufficient resources to deliver regular and high-quality formative feedback, especially in large classrooms or online learning environments, despite its essential significance across nearly all academic fields. In this context, diverse AI-driven systems, including Natural Language Processing (NLP) [8], [9], intelligent tutoring systems [10], [11], and autocorrect tools [12], [13], have been employed to facilitate the effective and efficient enhancement of argumentation skills.

There is a growing body of research on the use of AI in formative assessment to help people improve their argumentation skills [14]. However, a thorough examination of global research trends, key contributors, and dominant themes at the intersection of AI, formative assessment, and the improvement of argumentation skills, particularly in higher education, has not yet been conducted. As a result, this study aims to conduct a bibliometric analysis to understand the evolving research landscape in this field. This study specifically seeks to address the following inquiries:

RQ1: How are the research trends related to the use of AI in formative assessments for the development of argumentation skills in higher education?

RQ2: In this field, who are the most important authors, institutions, and countries?

RQ3: What are the most common keywords and study themes, and how do scientific networks come together?

RQ4: What research gaps and prospective research directions may be discerned from the existing literature?

Using a bibliometric approach, this study is likely to make a genuine difference by giving people more information and opening up new avenues for more focused and impactful research.

2. Methodology

This study employs bibliometric analysis, a statistical technique utilizing a quantitative framework to systematically examine research trends and their developmental trajectories [15].

This methodology serves as an effective instrument for analyzing the historical evolution of a discipline, the organization of scientific entities, patterns of information transmission, journal impact, and enduring citation trends [16]. Moreover, bibliometric analysis facilitates a more thorough comprehension by visualizing the research environment within a specific subject [17]. Consequently, this methodology aims to furnish comprehensive insights into the application of AI in formative assessment to enhance the cultivation of argumentation skills in higher education.

The bibliographic information for this study came from the Scopus database. Scopus was chosen due to several robust methodological factors, including its comprehensive coverage of scientific literature [18], [19], its superior curation quality [20], and its facilitation of metadata export necessary for bibliometric analysis [21]. These factors not only guarantee the precision and comprehensiveness of data but also provide a more methodical and thorough analysis.

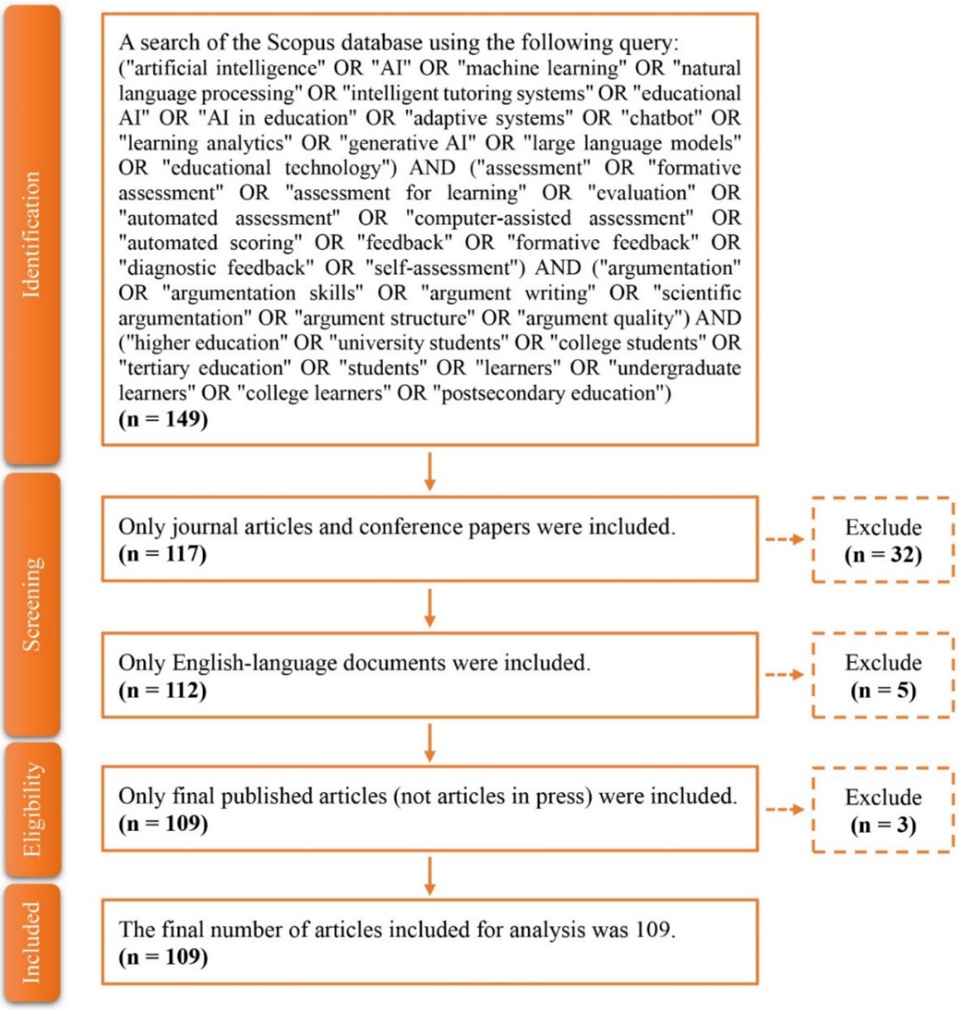

Figure 1. The steps of article selection

This study is confined to a singular database, Scopus, which may result in the exclusion of certain publications from alternative databases, such as Web of Science or IEEE Xplore, in the analysis. Additionally, the choice of keywords and how they are put together has a big effect on how broad and relevant the data coverage is. Therefore, the search strategy was carefully and deliberately planned to reduce bias and ensure that all relevant papers were included.

The data search and collecting for this study took place in August 2025. The PRISMA principles guided this method, ensuring an open and thorough screening process [22]. Figure 1 shows the different steps in the selection process.

This study used a mix of VOSviewer and Biblioshiny software to look at bibliographic data that was saved in .csv or .ris format. VOSviewer was selected for its capacity to generate network-based visualizations that facilitate intuitive comprehension of links among objects, including co-authorship, co-citation, and keyword co-occurrence [23]. Biblioshiny, a graphical interface of the bibliometrix package in R, was utilized for descriptive bibliometric analysis, encompassing the identification of annual publication patterns, main literature sources, and the most prolific authors or institutions [24]. Combining these two software applications allows for a more detailed examination of the scientific landscape within the examined field.

3. Results

The findings of this study indicate that research trends on the use of AI in formative assessment for the development of argumentation skills have continuously evolved from the early 1990s to 2025. Other findings highlight the authors, institutions, and countries that dominate research in this field. In addition, the study identifies key terms and major research themes, as well as the scholarly networks that have emerged. Based on these findings, research gaps and future research directions can be mapped. The detailed results of the study are presented as follows.

3.1. Research Trends

According to Scopus statistics, the main research on using AI in formative assessments to help college students improve their argumentation skills has grown a lot in the last few years. The average number of citations per document shows that this field has a big impact. Table 1 displays the characteristics of the 109 documents examined.

Table 1. Main information

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	1992:2025
Sources (Journals, Books, etc.)	51
Documents	109
Annual Growth Rate %	6.5
Document Average Age	6.94
Average citations per doc	24.39
References	925
DOCUMENT CONTENTS	
Keywords Plus (ID)	586
Author's Keywords (DE)	784
AUTHORS	
Authors	337
Authors of single-authored docs	9
AUTHORS COLLABORATION	
Single-authored docs	9
Co-Authors per Doc	3.83
International co-authorships %	27.52
DOCUMENT TYPES	
Article	54
Conference paper	55

Research in this field commenced in 1992 and continues to expand at an annual growth of 6.5%, indicating scholarly interest in the application of AI in formative evaluation to enhance argumentation skills in higher education. This growth is also a sign of the global push for new ways to teach using digital technologies. Figure 2 shows how publication trends have changed over time. It shows how productivity has gone up and down and how research contributions have gone up and down at certain times, probably because of changes in technology or educational regulations.

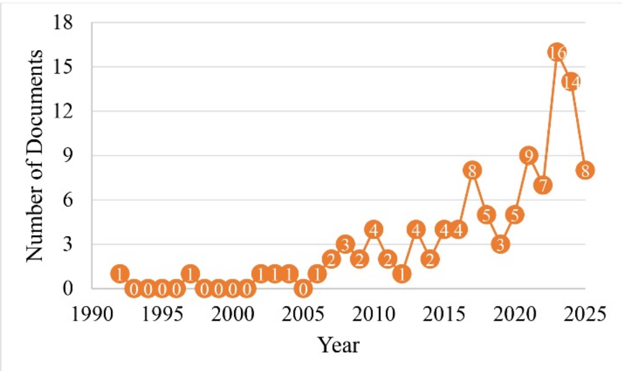

Figure 2. Yearly scientific output

Figure 2 depicts the trend in yearly publications on the subject from the early 1990s to 2025. Publications were few and irregular between 1992 and 2005, which reflected the field's infancy.

The introduction of technologies like machine learning and NLP in education, which sparked interest in applying AI to formative assessment, led to a moderate increase in publications from 2006 to 2016, with an average of two to six per year. Between 2017 and 2023, there was a notable increase, reaching a peak of more than 15 publications in 2023. The topic became a global research focus as a result of this surge, which also coincided with the emergence of learning analytics, AI-driven feedback systems, the widespread shift to online learning following COVID-19, cross-disciplinary collaborations, and growing interest in using Large Language Models (LLMs) in assessment. The number of publications decreased from 2023 to 2024 – 2025, but it was still higher than it was before 2020. The natural cycles of academic publishing, like peer review delays or a shift in research interest toward new fields like generative AI, could be the cause of this decline. However, the general upward trend indicates that the subject remains important and has a lot of room for innovation in higher education.

3.2. The Most Influential Authors, Institutions, and Countries

Finding the most important authors is key to comprehending the AI research landscape for a formative assessment of argumentation skills in higher education. The mapping shows where academic work is being done, important people, and possible collaborations that could change the course of future research [25]. Figure 3 shows a list of the most important writers based on bibliometric analysis. This format is a good way to look at their work, where they work, and their academic networks.

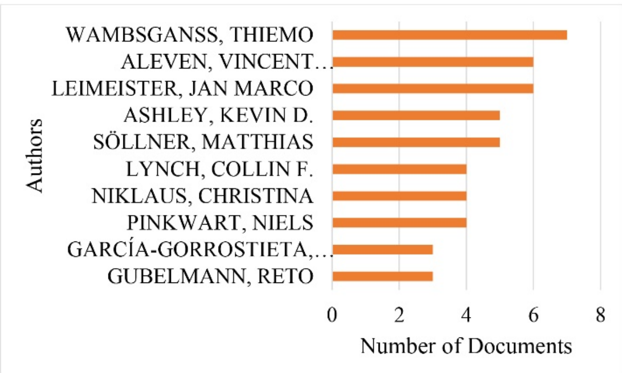

Figure 3. The most important authors

Thiemo Wambsganss is the author with the most publications (7 documents), followed by Vincent A.W.M.M. Aleven and Jan Marco Leimeister (6 documents), and Kevin D. Ashley and Matthias Söllner (5 documents). The other writers have written three to four books.

This distribution shows the main writers who are responsible for the growth of this subject, in line with Lotka's law, which says that most research productivity comes from a small number of authors with a lot of output [26], [27].

Figure 4 depicts author productivity in accordance with Lotka's law, demonstrating that the majority of authors only contribute one publication, with the number of authors drastically declining as the number of publications rises. The dashed line represents Lotka's theoretical curve, while the solid line displays actual data [28]. The close fit between the two indicates a consistent distribution pattern, despite the presence of a few fewer highly productive authors than expected. Even though core authors are still essential to the continuity and progress of research, the data emphasizes the value of sporadic contributors. Authors like Vincent A.W.M.M. Aleven and Thiemo Wambsganss are well known as thought leaders who support the advancement of the field and maintain the caliber of research. However, if new researchers with different viewpoints are not included, their dominance may also restrict methodological diversity and impede innovation. Therefore, cooperation between senior scholars and early-career researchers is crucial to ensure the field stays dynamic and changes over time. Figure 5 displays the author collaboration map.

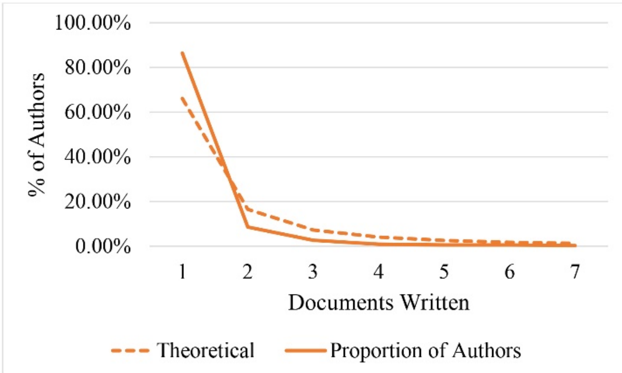

Figure 4. Author productivity through Lotka's law

With over ten groups, Figure 5 shows a collaborative network of authors. Vincent A.W.M.M. Aleven, Kevin D. Ashley, and Niels Pinkwart lead the core cluster (red), which has more than five connections each, indicating ongoing collaboration on related topics [29]. Large nodes represent authors with high publication and citation counts [30], [31]. With three to four connections and a focus on different themes, the blue cluster, which includes Thiemo Wambsganss and Matthias Söllner, and the green cluster, which includes Jesús Miguel García-Gorrostieta and Samuel Gonzalez-Lopes, remain only marginally connected to the rest of the network. Perhaps as a result of their affiliation with particular institutions or geographical areas, smaller clusters and lone authors, like Xiaoming Zhai, exhibit less cooperation.

Global research participation across a range of scientific affiliations (Figure 6) and backgrounds is reflected in the network's names, which come from Asia, Europe, and Latin America. However, less than 10% of all edges are inter-cluster connections, indicating a large area for improvement in international cooperation [32].

The red cluster could be a major hub for author connections, idea sharing, and network growth. Participating in cross-national or interdisciplinary projects with less connected authors may result in more citations and the introduction of fresh viewpoints.

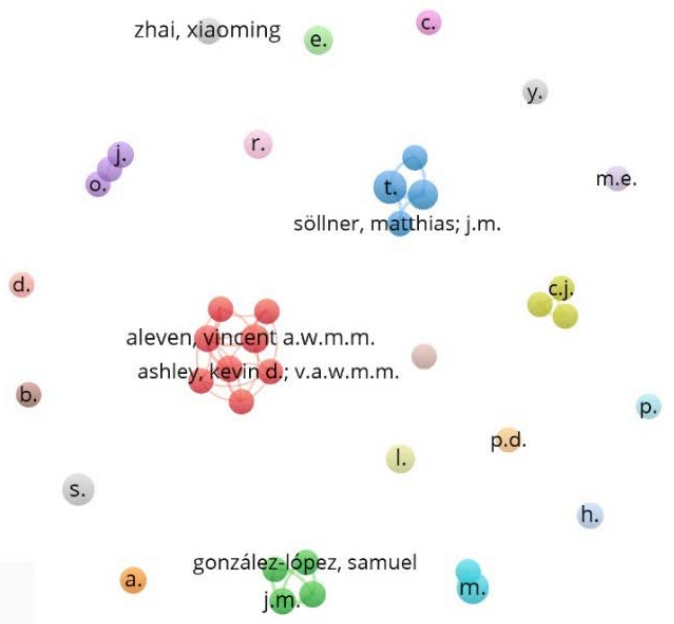

Figure 5. Map of collaboration between authors

Figure 6 shows that the University of St. Gallen is the world's top research institution in this area, with 32 publications. Following it are the University of Pittsburgh with eleven publications, National Dong Hwa University with fifteen, and The Open University with ten, all of which have made noteworthy contributions to their respective regions. The diversity of viewpoints and approaches in the field is supported by the moderate but significant contributions (7–8 publications) made by organizations like the School of Computer Science, Northern Illinois University, and the University of California. Strong interest from the technological and applied science sectors is further evidenced by research-focused organizations such as the Instituto Nacional de Astrofísica Óptica y Electrónica, the Erik Jonsson School of Engineering and Computer Science, and the Concord Consortium. A wide range of general universities, specialized engineering schools, and focused research centers are making significant contributions, according to this distribution pattern, even though a small number of universities dominate in terms of volume. This suggests a significant chance for greater inter-institutional collaboration, which may lead to future innovation and wider research applications.

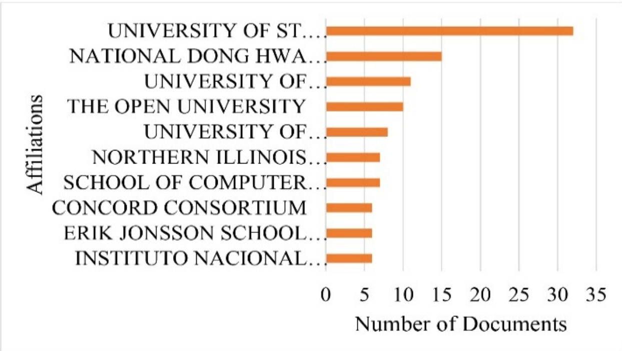

Figure 6. The most important affiliation

When looking at co-authorship at the affiliate level, Figure 7 shows that there are only two small collaborative clusters connected and that research in this area is still concentrated within a few major universities. The University of St. Gallen and Universität Kassel, which both contribute significantly but have small networks of partners, are in charge of the green cluster. Under the direction of the School of Computer Science, the red cluster facilitates transatlantic cooperation between the USA and Europe by acting as a vital link between the University of Pittsburgh and Technische Universität Clausthal.

On the other hand, organizations like the Educational Testing Service, Northern Illinois University, the University of Georgia, The Open University, and the Instituto Nacional de Astrofísica Óptica y Electrónica typically function independently, having close internal relationships but few external alliances. These institutions have not yet fully utilized international collaboration networks, despite their strength in research, and this fragmentation can impede cross-national and interdisciplinary knowledge exchange [33]. Therefore, it is necessary to strengthen inter-institutional collaboration to improve research quality, increase the number of citations, and advance science more quickly.

Figure 7 also shows that research activities are mostly restricted to a small number of powerful organizations, which unequally distribute resources and restrict international participation.

Institutional collaboration patterns remain weakly connected. This emphasizes how important it is to foster international collaborations to create research ecosystems that are more resilient, diverse, and inclusive. Promoting wider cooperation can broaden methodological and theoretical viewpoints while lessening institutional domination. Analyzing contributions by nation of origin is also crucial for obtaining a better grasp of global research dynamics. Figure 8 illustrates this more comprehensive view by highlighting the geographic distribution of publications. It also shows how national research capacity, international collaboration networks, and institutional quality affect scientific productivity globally.

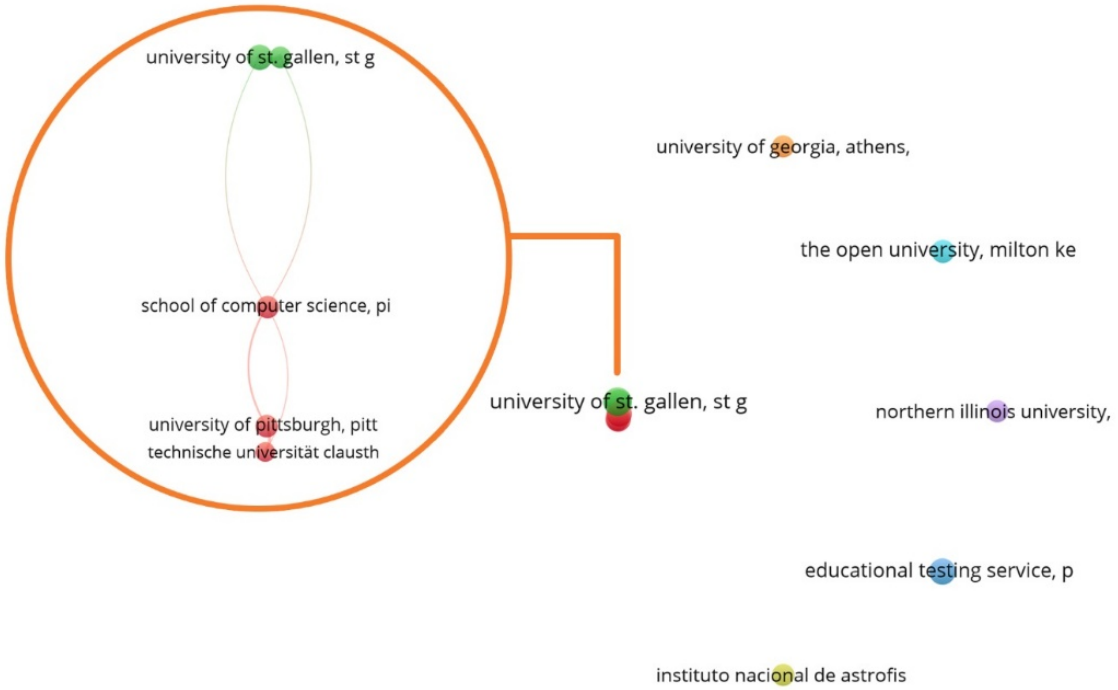

Figure 7. A map of how affiliates work together

Figure 8 illustrates the global distribution of research contributions. The colors show how many publications there are in each country. The USA is the leader in scientific research, followed by Western European countries like the UK, Germany, and France, and Asian countries like China, Japan, and South Korea. A well-developed research ecosystem, a lot of financing, a lot of international cooperation, and good technology infrastructure all support this advantage.

Developing nations, notwithstanding their diminished contributions, exhibit strategic potential through localized perspectives. The disparity in research between developed and developing nations creates the potential to enhance cross-border collaboration, improve methodology, amplify impact, and advocate for the equitable allocation of global scientific contributions.

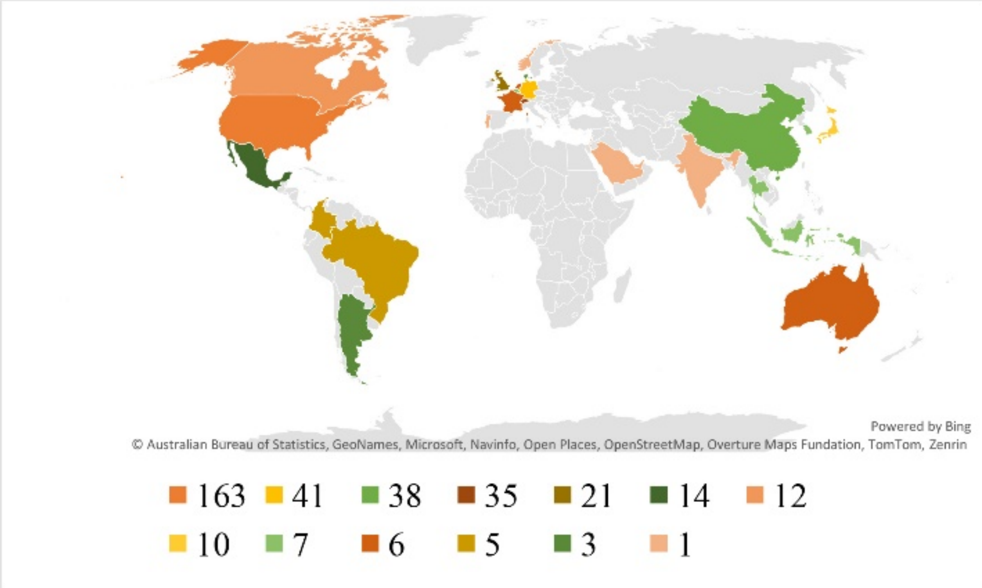

Figure 8. The most important countries

Figure 9 presents a co-authorship analysis at the national level, identifying three main clusters. Strong cooperation between the USA, Germany, and Switzerland is demonstrated by the first cluster (blue), with the USA acting as a major hub for international research, especially with European partners like Germany. Despite being smaller than the USA cluster, the second cluster (red), which is headed by the UK and consists of Australia, France, Japan, and the Netherlands, serves as a vital regional bridge. Although it has few links to international research centers, the third cluster (green), which consists of China, Singapore, and South Korea, indicates potential for increased international integration and demonstrates the growing East Asian cooperation centered around China.

Outside of these major clusters are nations like Canada, Mexico, Taiwan, Belgium, Indonesia, and others that frequently publish independently and have few connections to international networks, which may lessen the visibility and impact of their research. These outlying countries could, however, improve cooperation with important centers such as the USA, UK, and China, promoting knowledge sharing, raising the caliber of research, and building a more diverse global research environment. Overall, the map emphasizes how the USA dominates global cooperation, whereas Europe and Asia form powerful but geographically dispersed groups. Increasing cross-cluster collaboration may address this fragmentation and foster a more cohesive international research community [34].

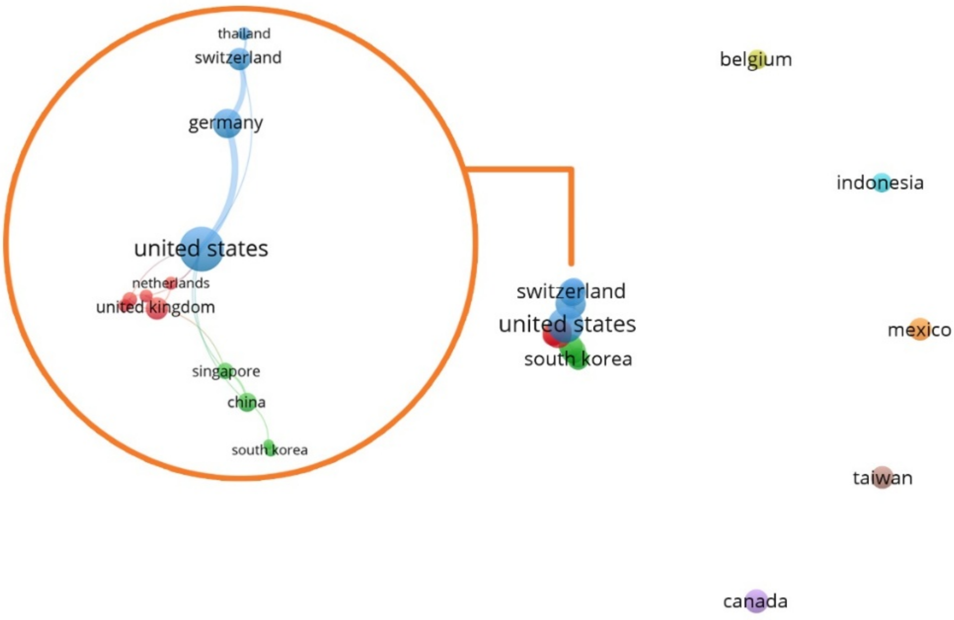

Figure 9. Map of collaboration between countries

3.3. Most Frequently Appearing Keywords and Research Themes, and Established Scientific Networks

Figure 10 illustrates a word cloud analysis that indicates the word "students" as the most important one. This means that the research topic is very student-centered. Other words that show a connection between academic skill development and the use of AI technology in education are "natural language processing systems", "education", "argumentation", "artificial intelligence", and "machine learning." This substantiates that the research is not exclusively educational but also utilizes advancements in computational linguistics.

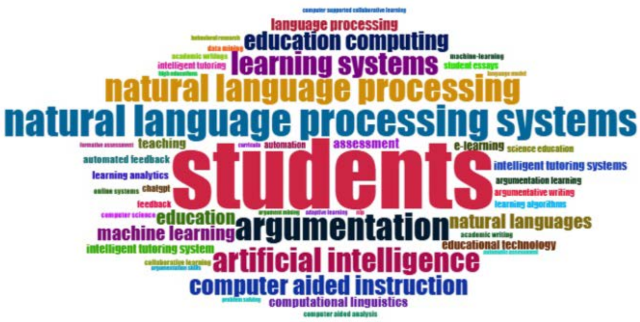

Figure 10. Word cloud

The results of the word cloud present a general idea of the research emphasis based on keyword frequency, but they do not yet explain how themes are related to each other or how concepts have changed over time.

For instance, the prevalence of the term "students" does not elucidate its particular context in AI, debate, or NLP. To gain a more profound comprehension, visual analysis techniques like thematic maps, which categorize and arrange subjects according to significance and scientific development, are essential [35], as illustrated in Figure 11.

Figure 11 displays a thematic map that highlights important keywords and themes within the overall research strategy. Argumentation learning, academic writing, learning analytics, collaborative learning, and argumentation skills are all well-developed and significant topics in the Motor Themes quadrant that act as domain bridges and drivers. Keywords such as "students", "natural language processing systems", and "argumentation" are also present in this quadrant, underscoring the crucial role of NLP technology in college-level argumentation instruction. Highly specialized but less generally applicable subjects like chromosomes, intelligent learning environments, and argumentation skills are found in the Niche Themes quadrant, which denotes developed research fields with little generalizability. The Basic Themes quadrant, on the other hand, contains pertinent but undeveloped topics like computer science, educational technology, intelligent tutoring systems, and e-learning, creating a solid foundation ready for more research.

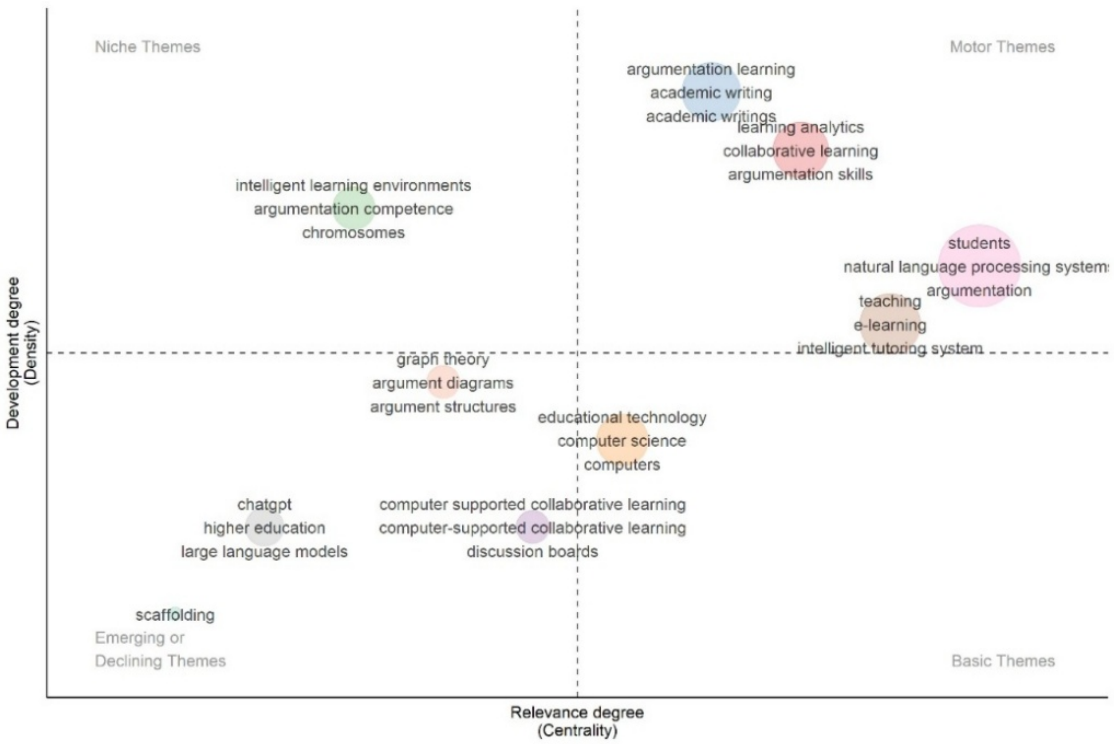

Figure 11. Thematic map

Newer subjects like ChatGPT, higher education, massive language models, and scaffolding are included in the Emerging or Declining Themes quadrant, which represents developing fields with substantial room for expansion. These themes are probably growing and might eventually turn into motor themes due to advancements in generative AI.

In general, research in technology-based learning and argumentation is advancing toward the integration of AI, collaboration, and learning analytics, with encouraging prospects for additional advancement in emerging domains. Keyword analysis reveals a scientific network, where each term represents a fundamental concept and its relation to others in the field. Figure 12 visualizes this network, highlighting key thematic areas, research trajectories, and patterns of conceptual links.

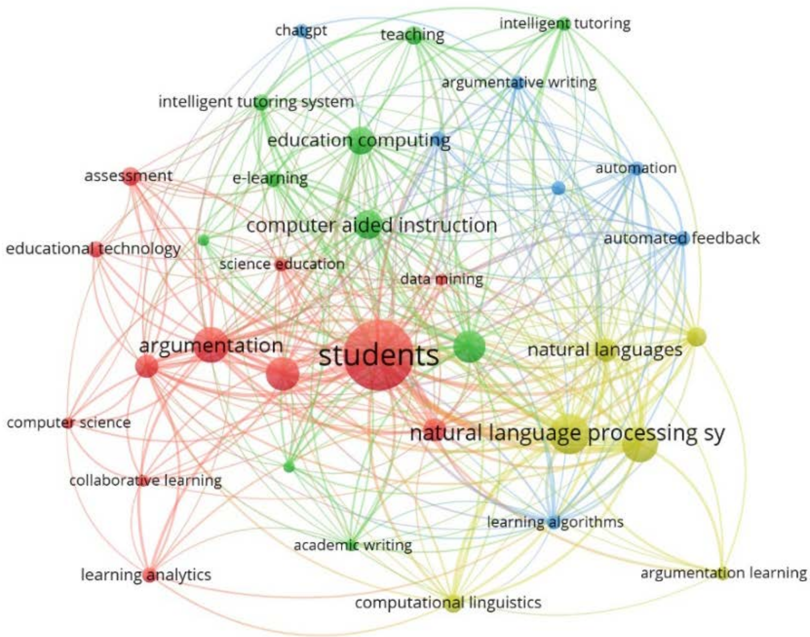

Figure 12. The network of keywords

The keyword network has four major clusters, as shown in Figure 12. The largest cluster, Cluster 1 (red), focuses on students and argumentation while examining AI and its use in argumentative learning in the classroom. It links technology to education by fusing pedagogical components like collaborative learning and educational technology with technical ones like data mining and machine learning. This cluster, however, runs the risk of technocentrism, in which the study of argumentation theory is subordinated to the use of technology. Cluster 2 (green) centers on e-learning and instructional systems, utilizing terms such as computer-aided instruction, e-learning, and intelligent tutoring. This cluster places a strong emphasis on digital tools for formal education, which increases student engagement but runs the risk of taking a didactic approach that puts content delivery ahead of student-centered argumentation.

With terms like ChatGPT, learning algorithms, and automated feedback, Cluster 3 (blue) focuses on argumentative writing and LLM-based feedback automation.

Although issues such as accuracy, context sensitivity, and potential biases persist, the use of AI-driven, scalable, and effective feedback enhances argumentation assessment. To strike a balance between technology and human pedagogical insight, future research is encouraged to combine AI with teacher-in-the-loop strategies. To create automated argumentation assessment systems, Cluster 4 (yellow) covers fundamental NLP and computational linguistics, including argument mining and deep learning. However, to progress the field, NLP must be more deeply integrated with argumentation theory due to its limitations in capturing rhetorical, semantic, and pragmatic nuances. Together, these clusters form an interdisciplinary framework: Cluster 4 provides the technical framework, Cluster 3 enhances it by automating feedback, Cluster 2 integrates these technologies into digital learning environments, and Cluster 1 prioritizes students and practical argumentative skills. This approach emphasizes that technological advancement is only significant when it is in line with efficient learning procedures and well-defined educational goals [36], [37].

3.4. Research Gaps and Prospective Research Avenues Identified from The Existing Literature

The bibliometric analysis depicted in Figure 12 identifies four primary clusters and a distinct historical trend in research concerning the application of AI for formative assessment and its facilitation of students' argumentation skill development in higher education. These results establish a crucial basis for pinpointing research gaps and devising future trajectories for study advancement.

First, the red cluster shows how closely related argumentation, assessment, learning analytics, and collaborative learning are. This group of studies primarily examines the application of AI in tests and educational technology. Nonetheless, AI-based assessment is infrequently associated with the enhancement of argumentation skills, as research often measures learning outcomes without evaluating the quality of reasoning or the dynamics of discussion. Future research must provide an AI-driven formative assessment framework that evaluates both the end outcome and the process of argumentation development within collaborative and inquiry-based learning contexts.

The second, green cluster is based on research into computer-aided instruction, intelligent tutoring systems, and academic writing. While smart tutoring systems represent a significant application of AI in education, their use extends to numerous other domains. Moreover, the relationship between academic writing and AI-based formative feedback remains relatively weak. This indicates that further research is necessary to develop AI-based writing assistants that support argumentative writing and to evaluate their effectiveness in real classrooms, rather than just in labs or simulations.

Third, the blue cluster shows study areas that are growing quickly: argumentative writing, automated feedback, ChatGPT, and student essays.

The focus of research is transitioning towards generative AI and automated evaluation; yet, investigations into validity, fairness, and the effects of algorithmic bias and cultural context are still insufficient. Concerns regarding pupils' dependence on AI are hardly examined. Consequently, subsequent research must focus on creating understandable AI feedback systems that adhere to educational norms and undertake longitudinal studies to evaluate their enduring effects on students' critical thinking and argumentative skills.

Fourth, the yellow cluster shows the research's methodological foundations, which are NLP and computational linguistics. NLP has made progress in finding argument structure, but it still only looks at surface-level things like claims, evidence, and reasoning. It does not look at things like emotions, rhetorical dimensions, or socio-cultural background very much. Furthermore, the dominance of English highlights gaps in multilingual contexts. Consequently, subsequent research must focus on the development of multilingual NLP models and the integration of linguistic aspects with pedagogical frameworks to guarantee that evaluations are both technically sound and educationally pertinent.

Figure 13 (overlay visualization) illustrates the time history of this study's development. The main things worked on in the first few years (2014–2017) were educational technology, collaborative learning, and learning analytics. From 2017 to 2019, the focus changed to intelligent tutoring systems and academic writing. This year was the start of using AI in learning. Research into NLP, automated feedback, and essay analysis grew considerably between 2019 and 2021. This episode shows how AI is becoming more and more important in text-based tests. Finally, the years 2021–2022 saw the rise of generative AI like ChatGPT, which brings with it new chances as well as moral, teaching, and methodological problems.

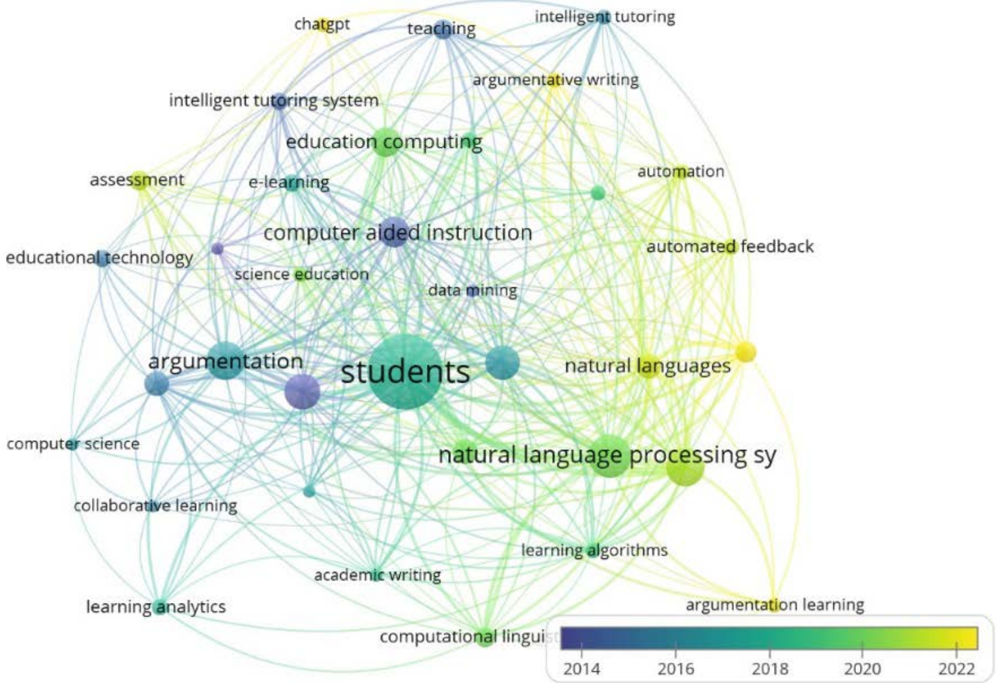

Figure 13. Overlay visualization

In general, research trends show a move away from generic AI applications and toward using text-based generative AI to help students improve their argumentation skills directly. To facilitate advancement, forthcoming research must: (1) Investigate the enduring effects of AI feedback on argumentative competencies; (2) Tackle bias and equity challenges within multilingual and cross-cultural settings; and (3) Formulate pedagogical frameworks that correspond with formative assessment and inquiry-driven education. These stages are very important to make sure that AI is not just an automated testing tool but also really helps students in higher education learn how to argue at a higher level.

4. Discussion

Research trends have shown steady growth since 1992, at a pace of 6.5% per year. This means that the research base is mature but not yet saturated. This continual expansion creates chances for new ideas, especially in the creation of AI-based evaluation systems that give automatic, tailored, and contextual feedback on how students argue. The implementation of deep learning technologies [38], [39], [40] and LLM [41], [42], [43] facilitates the incorporation of real-time argumentation assessments into Learning Management Systems (LMS), necessitating interdisciplinary research across domains such as education, linguistics, AI ethics, and culture.

In the future, research should concentrate on (1) Creating inclusive multilingual benchmark datasets; (2) Designing adaptive assessments customized to student profiles; and (3) Fostering AI-lecturer collaboration in hybrid assessment systems that integrate machine analytics with pedagogical insight. Following this path, research will not only be able to continue, but it will also lead to wiser, more ethical, and more effective ways to judge arguments.

The triadic structure of author, institution, and country in this field creates a centralized hierarchy that makes sure research continues but does not allow for many different contexts. Knowledge production is concentrated among a small group of prolific authors, such as Thiemo Wambsganss, Vincent A.W.M.M. Aleven, Jan Marco Leimeister, Kevin D. Ashley, and Matthias Söllner, aligning with Lotka’s Law, which states that most authors publish only once while a select few drive the field forward [44]. The University of St. Gallen, National Dong Hwa University, the University of Pittsburgh, and The Open University are all important institutions. The Concord Consortium is an example of a technical contributor. The USA is in charge at the national level, with Germany and Switzerland as important partners. The UK is in charge of the European cluster, and East Asia (China, Singapore, and South Korea) is in charge of its cluster. Belgium, Indonesia, Mexico, Taiwan, and Canada are still on the outside. This setup lets central players set research standards and control strategic alliances (like the USA–mainland Europe axis), but less than 10% of collaborations cross cluster boundaries, which shows that the global network is broken up.

This setup makes it easier to keep things going and quickly adopt new methods, but it also makes path dependence and center-periphery divides stronger, which limits methodological diversity and context sensitivity. Three policy implications come to light from a strategic perspective. First, planned cross-institutional consortia should be created to link core institutions with those that are more isolated. This will encourage collaboration between clusters and the use of different methods. Second, researchers should be able to move more easily between countries so that the scientific strengths of the USA and Europe can be combined with the cultural contexts of Asia and the Global South. Third, funding and curation systems should support different methods, like including local contexts, using non-traditional assessments, and doing cross-country replication studies. This will assure that collaboration between authors, institutions, and countries not only increases scientific output but also leads to more innovation and a bigger impact on society.

The keyword network map (Figure 12) shows a planned path toward innovation by showing how research has moved through four main phases, each of which is linked to a specific theme. Phase 1 (cluster 4 – yellow) establishes the methodological foundation by enhancing NLP and computational linguistics, concentrating on the refinement of language processing and argumentation learning algorithms [45], [46], [47]. Even though deep learning and argument mining have made a lot of progress, it is still challenging to understand the semantics and rhetorical structures of argumentative texts [48].

Phase 2 (cluster 3 – blue) signifies a transition towards educational applications, particularly in the automation of argumentative feedback utilizing AI and LLMs such as ChatGPT [12], [49], [50]. This stage allows for scalable, individualized feedback, but it also raises issues like automation bias and difficulties in understanding subtle differences in discourse [51].

Phase 3 (cluster 2 – green) adds automated feedback to larger instructional systems, like computer-aided instruction and intelligent tutoring systems, moving closer to personalized AI-supported learning [52]. This makes things more efficient, but it could also make the learning process too structured and take away some of the students' freedom [53]. Phase 4 (cluster 1 – red) signifies complete pedagogical implementation, wherein technology facilitates authentic classroom learning to augment students' critical, collaborative, and argumentative thinking [54], [55], [56]. At this point, educational success depends on making AI tools work well with teaching methods like formative assessment and collaborative learning [57]. This roadmap shows a clear progression, from basic technologies to classroom integration, and it shows how AI and NLP could help people learn how to argue in the 21st century.

Four main groups in the keyword network (Figure 12) show big holes in the research and possible future research paths. Each cluster signifies an underexplored thematic focus, thereby creating avenues for further research advancement. Table 2 displays the identified research gaps and opportunities. It also shows the contributions of each cluster and suggestions for more research on the topic.

Table 2. Summary of clusters, research gaps, and future research directions

Clusters	Research Focus	Research Gaps	Future Research Directions
Red	Argumentation, assessment, collaborative learning, learning analytics	The emphasis continues to be on learning outcomes, with minimal attention given to assessing the quality of the argumentative process. The combination of AI-based tests and collaborative learning is still not very common.	Creating AI-based formative assessments that evaluate the construction of arguments (claims, evidence, reasoning) within a collaborative, inquiry-driven framework.
Green	Intelligent tutoring systems, academic writing, e-learning	Intelligent tutoring systems remain broad in scope and have not been explicitly designed to enhance argumentation skills. There has not been much research on how AI feedback affects academic writing and argumentation.	Creating an AI-based tutor/writing system that helps with argumentative writing and testing it in a real classroom.
Blue	Argumentative writing, automated feedback, ChatGPT, student essays	The validity and fairness of automated feedback are still not well studied. The risks of bias, cultural context, and dependence on AI have not been comprehensively examined.	Creating AI feedback that is clear and fits with how people learn. Doing a long-term study to see how using ChatGPT affects critical thinking skills over time.
Yellow	Natural Language Processing, computational linguistics	NLP is more concerned with the surface structure of arguments than with rhetoric, emotions, and social and cultural factors. It is mostly in English, but not in other languages.	Creating multilingual NLP and models that take into account rhetorical and socio-cultural aspects. Combining NLP with a teaching framework for formative assessment.

Figure 13 shows an overlay visualization that shows how research on AI-supported formative assessment to improve students' argumentation skills is spread out over many fields, with each one focusing on a different technological aspect. There has been a lot of focus on NLP, intelligent tutoring systems, and automated feedback, but pedagogical integration, especially context-sensitive formative assessment that aims to improve the quality of arguments, is still not well understood.

Table 2 shows important gaps in research, such as not measuring the quality of arguments enough, not using AI enough in multilingual and collaborative settings, and not paying enough attention to the validity and fairness of automated feedback. These results highlight the necessity for an interdisciplinary framework that integrates AI, formative assessment, and argumentative learning theory. Future research should focus on three key priorities: (1) Creating AI systems that adhere to pedagogical principles for providing formative, scaffolding-based feedback; (2) Implementing AI in various contexts, collaborative, multilingual, and cross-cultural, to enhance global higher education; and (3) Executing longitudinal studies to evaluate AI's enduring influence on students' argumentation, writing skills, and readiness for 21st-century challenges.

5. Conclusion

Research on the utilization of AI in formative assessments for the enhancement of argumentation skills in higher education has demonstrated a consistent upward trajectory since the 1990s, with significant contributions from select authors, institutions, and key nations, including the USA, Germany, and Switzerland. This concentration enhances research continuity and sets global methodological standards; however, it also risks homogenizing perspectives that are predominantly focused on developed country contexts. Keyword network analysis shows a clear path for research: it starts with the basics of NLP, then moves on to automating feedback on arguments, adding it to digital teaching systems, and finally using it in real-life teaching.

Even though there has been a lot of progress, there are still some important research gaps. For example, the argumentation process dimension has not been studied enough, and it has not been used in multilingual and collaborative settings. Also, automated feedback is challenging to make valid and fair. Future research should focus on the integration of AI with pedagogical principles, the enhancement of collaboration across institutions and nations, and the adoption of an interdisciplinary approach, ensuring that the technology is not only technically sophisticated but also effective in enhancing students' critical thinking and argumentative skills.

Acknowledgements

Universitas Negeri Surabaya (Unesa) provided funding for this study through the 2025 Non-State Budget (Non-APBN) program for International Collaborative Research with the top 100 universities worldwide, in accordance with the Rector's Decree of Unesa Number 506/UN38/HK/2025. The authors extend their sincere gratitude to Unesa for her kind assistance and facilitation, which made this study possible.

References:

[1]. Bewersdorff, A., et al. (2023). Assessing student errors in experimentation using artificial intelligence and large language models: A comparative study with human raters. *Computers and Education: Artificial Intelligence*, *5*, 100177. Doi: 10.1016/j.caeai.2023.100177

[2]. Swiecki, Z., et al. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *3*, 100075. Doi: 10.1016/j.caeai.2022.100075

[3]. Yao, Y., et al. (2025). Secondary school English teachers' application of artificial intelligence-guided chatbot in the provision of feedback on student writing: An activity theory perspective. *Journal of Second Language Writing*, *67*, 101179. Doi: 10.1016/j.jslw.2025.101179

[4]. Liebenow, L. W., et al. (2025). Self-Assessment Accuracy in the Age of Artificial Intelligence: Differential Effects of LLM-Generated Feedback. *Computers & Education*, 105385. Doi: 10.1016/j.compedu.2025.105385

[5]. Meyer, J., et al. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, *6*, 100199. Doi: 10.1016/j.caeai.2023.100199

[6]. Giri, V., & Paily, M. U. (2020). Effect of scientific argumentation on the development of critical thinking. *Science & Education*, *29*(3), 673-690. Doi: 10.1007/s11191-020-00120-y

[7]. Fauziah, A. N. M., et al. (2024). Relationship Between Critical Thinking and Scientific Argumentation in Science Learning. *Jurnal Pendidikan IPA Indonesia*, *13*(2). Doi: 10.15294/pwqxcp96

[8]. Guo, K., et al. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers & Education*, *203*, 104862. Doi: 10.1016/j.compedu.2023.104862

[9]. Wambsganss, T., Guggisberg, S., Söllner, M. (2021). ArgueBot: A Conversational Agent for Adaptive Argumentation Feedback. *Innovation Through Information Systems. WI 2021. Lecture Notes in Information Systems and Organisation*, *47*. Doi: 10.1007/978-3-030-86797-3_18

[10]. Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, 100752. Doi: 10.1016/j.asw.2023.100752

[11]. Li, H., et al. (2025). The influence of GenAI on the effectiveness of argumentative writing in higher education: Evidence from a quasi-experimental study in China. *Journal of Asian Public Policy*, *18*(2), 405-430. Doi: 10.1080/17516234.2024.2363128

[12]. Wambsganss, T., Janson, A., & Leimeister, J. M. (2022). Enhancing argumentative writing with automated feedback and social comparison nudging. *Computers & Education*, *191*, 104644. Doi: 10.1016/j.compedu.2022.104644

[13]. Zhu, M., Liu, O. L., & Lee, H. S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education*, *143*, 103668. Doi: 10.1016/j.compedu.2019.103668

[14]. Memarian, B., & Doleck, T. (2024). A review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100040. Doi: 10.1016/j.chbah.2023.100040

[15]. Prahani, B. K., et al. (2024). Mapping research on scientific creativity: A bibliometric review of the literature in the last 20 years. *Thinking Skills and Creativity*, *52*, 101495. Doi: 10.1016/j.tsc.2024.101495

[16]. Muchemwa, M., Sodi, T., & Themane, M. (2024). A bibliometric analysis of mental health research in places of higher learning in sub-Saharan Africa. *Social Sciences & Humanities Open*, *10*, 100950. Doi: 10.1016/j.ssaho.2024.100950

[17]. Amiruddin, M. Z. B., et al. (2025). Scientific mapping and trend of conceptual change: A bibliometric analysis. *Social Sciences & Humanities Open*, *11*, 101208. Doi: 10.1016/j.ssaho.2024.101208

[18]. Hafiar, H., et al. (2024). Scoring big: A bibliometric analysis of the intersection between sports and public relations. *Cogent Social Sciences*, *10*(1), 2297706. Doi: 10.1080/23311886.2023.2297706

[19]. Challco, G. C., et al. (2024). Blended learning and media centers: A bibliometric analysis. *Social Sciences & Humanities Open*, *10*, 100919. Doi: 10.1016/j.ssaho.2024.100919

[20]. Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*, *9*(1), 12. Doi: 10.3390/publications9010012

[21]. Hamid, S., & Huda, M. N. (2025). Mapping the landscape of government data breaches: A bibliometric analysis of literature from 2006 to 2023. *Social Sciences & Humanities Open*, *11*, 101234. Doi: 10.1016/j.ssaho.2024.101234

[22]. Page, M. J., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 89. Doi: 10.1186/s13643-021-01626-4

[23]. Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *scientometrics*, *84*(2), 523-538. Doi: 10.1007/s11192-009-0146-3

[24]. Quoc Bui, D., et al. (2021). Two decades of corruption research in ASEAN: A bibliometrics analysis in Scopus database (2000–2020). *Cogent Social Sciences*, *7*(1). Doi: 10.1080/23311886.2021.2006520

[25]. Li, Y., & Zheng, X. (2025). A literature review of intercultural emotion studies from 1987 to 2023: Theoretical approaches, hot topics and future directions. *International Journal of Intercultural Relations*, *106*, 102179. Doi: 10.1016/j.ijintrel.2025.102179

[26]. Amarathunga, B. (2024). ChatGPT in education: unveiling frontiers and future directions through systematic literature review and bibliometric analysis. *Asian education and development studies*, *13*(5), 412-431. Doi: 10.1108/AEDS-05-2024-0101

[27]. Egghe, L. (1993). Consequences of Lotka's law in the case of fractional counting of authorship and of first author counts. *Mathematical and Computer Modelling*, *18*(9), 63-77. Doi: 10.1016/0895-7177(93)90143-M

[28]. Malcai, O., et al. (2002). Theoretical analysis and simulations of the generalized Lotka-Volterra model. *Physical Review E*, *66*(3), 031102. Doi: 10.1103/PhysRevE.66.031102

[29]. Hoorsun, A., et al. (2024). Insights and research trends of dust and cleaning in solar energy: A bibliometric review study. In *2024 9th International Conference on Technology and Energy Management (ICTEM)*, 1-5. Doi: 10.1109/ICTEM60690.2024.10632021

[30]. Yi, W., Zhao, Y., & Wang, W. (2025). Four decades of research on multiword expressions: A bibliometric analysis. *Applied Linguistics*. Doi: 10.1093/applin/amae085

[31]. Sun, L., et al. (2024). Application of Artificial Intelligence in the diagnosis and treatment of colorectal cancer: a bibliometric analysis, 2004–2023. *Frontiers in Oncology*, *14*, 1424044. Doi: 10.3389/fonc.2024.1424044

[32]. Hu, Y., et al. (2025). Bibliometric analysis of coating protection from 2015 to 2025. *Coatings*, *15*(6), 686. Doi: 10.3390/coatings15060686

[33]. Falcone, P. M., & Tutore, I. (2025). Mapping the nexus: A bibliometric analysis and social network analysis of transformative innovation policies and sustainable development goals. *Business Strategy and the Environment*, *34*(2), 2423–2435. Doi: 10.1002/bse.4104

[34]. Wan, Y., et al. (2025). Knowledge mapping of ultrasound technology and triple-negative breast cancer: A visual and bibliometric analysis. *Discover Oncology*, *16*(1), 1248. Doi: 10.1007/s12672-025-03063-4

[35]. Díaz, F., Liza, R., & Cerna, N. (2025). Mapping the Landscape of Blockchain for Transparent and Sustainable Supply Chains: A Bibliometric and Thematic Analysis. *Logistics*, *9*(3), 86. Doi: 10.3390/logistics9030086

[36]. Govindasamy, T. (2001). Successful implementation of e-learning: Pedagogical considerations. *The internet and higher education*, *4*(3-4), 287-299. Doi: 10.1016/S1096-7516(01)00071-9

[37]. Chugh, R., et al. (2023). Implementing educational technology in higher education institutions: A review of technologies, stakeholder perceptions, frameworks and metrics. *Education and Information Technologies*, *28*(12), 16403–16429. Doi: 10.1007/s10639-023-11846-x

[38]. Martin, P. P., et al. (2024). Applying machine learning for the analysis of student argumentation in chemistry: Applying machine learning for the analysis of student argumentation in chemistry. *Journal of Research in Science Teaching*, *61*(8), 1757–1792. Doi: 10.1002/tea.21903

[39]. Kang, K., & Sinnott, R. O. (2018). Improving online argumentation through deep learning. *Computational Science and Its Applications – ICCSA 2018*, 376–391. Doi: 10.1007/978-3-319-95162-1_26

[40]. Wambsganss, T., et al. (2020). AL: An adaptive learning support system for argumentation skills. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1-14. Doi: 10.1145/3313831.3376732

[41]. de la Puente, M., et al. (2024). Investigating the use of ChatGPT as a tool for enhancing critical thinking and argumentation skills in international relations debates among undergraduate students. *Smart Learning Environments*, *11*(1), 55. Doi: 10.1186/s40561-024-00347-0

[42]. de la Puente Pacheco, M. A., et al. (2025). Enhancing critical thinking and argumentation skills in Colombian undergraduate diplomacy students: ChatGPT-assisted and traditional debate methods. *Journal of Political Science Education*, 1–11. Doi: 10.1080/15512169.2025.2449936

[43]. Yang, K. B., Nam, S., Huang, Y., & Wood, S. (2024). Rhetor: Providing LLM-based feedback for students' argumentative essays. *Technology Enhanced Learning for Inclusive and Equitable Quality Education*, 201–205. Springer.

[44]. dell'Isola, F. (2024). Bibliometrics, Scientometrics and the Impact of Publishor Perish Culture on Continuum Mechanics. *The Principle of Virtual Work in the Sources of Mechanical Sciences: A Reappraisal of Pseudo-Aristotelian Mechanica Problemata and Germain's Article on Second Gradient Continuum Mechanics*, 187-253. Springer Nature Switzerland. Doi: 10.1007/978-3-031-74967-4_4

[45]. Manning, C. D. (2015). Computational linguistics and deep learning. *Computational Linguistics*, *41*(4), 701-707. Doi: 10.1162/COLI_a_00239

[46]. Bjerva, J. (2023). The role of typological feature prediction in NLP and linguistics. *Computational Linguistics*, *50*(2), 781-794. Doi: 10.1162/coli_a_00498

[47]. Tsujii, J. I. (2021). Natural language processing and computational linguistics. *Computational Linguistics*, *47*(4), 707-727. Doi: 10.1162/coli_a_00420

[48]. Hinton, M., & Wagemans, J. H. (2023). How persuasive is AI-generated argumentation? An analysis of the quality of an argumentative text produced by the GPT-3 AI text generator. *Argument & Computation*, *14*(1), 59-74. Doi: 10.3233/AAC-210026

[49]. Chen, X., et al. (2025). Effects of ChatGPT and argument map (AM)-supported online argumentation on college students' critical thinking skills and perceptions. *Education and Information Technologies*, *30*(12), 17623–17658. Doi: 10.1007/s10639-025-13471-2

[50]. Darmawansah, D., Rachman, D., & Febiyani, F. (2024). Collaborative Argument Map-Based ChatGPT to Enhance EFL Students' Argumentation Complexity and Group Efficacy. *Taiwan International ESP Journal*, *15*(2), 97-124. Doi: 10.6706/TIESPJ.202412_15(2).0004

[51]. Strauß, S. (2021). Deep automation bias: how to tackle a wicked problem of AI?. *Big Data and Cognitive Computing*, *5*(2), 18. Doi: 10.3390/bdcc5020018

[52]. Tamaki, K. Y., & Ishii, K. (2023). Automation in education, training, and learning systems. *Springer Handbook of Automation*, 1349-1370. Springer International Publishing. Doi: 10.1007/978-3-030-96729-1_63

[53]. Lawson McLean, A. (2024). Constructing knowledge: the role of AI in medical learning. *Journal of the American Medical Informatics Association*, *31*(8), 1797-1798. Doi: 10.1093/jamia/ocae124

[54]. Ruwe, T., & Mayweg-Paus, E. (2023). "Your argumentation is good", says the AI vs humans–The role of feedback providers and personalised language for feedback effectiveness. *Computers and Education: Artificial Intelligence*, *5*, 100189. Doi: 10.1016/j.caeai.2023.100189

[55]. Hackl, V., & Müller, C. (2022). AI-Supported Acquisition of Argumentation Skills: Use Case 'The Argueniser'. *International Conference on Design, Learning, and Innovation*, 48-54. Springer Nature Switzerland. Doi: 10.1007/978-3-031-31392-9_4

[56]. Wambsganss, T., et al. (2025). Improving students' argumentation skills using dynamic machine-learning–based modeling. *Information Systems Research*, *36*(1), 474-507. Doi: 10.1287/isre.2021.0615

[57]. Banihashem, S. K., Gašević, D., & Noroozi, O. (2025). A Critical Review of Using Learning Analytics for Formative Assessment: Progress, Pitfalls and Path Forward. *Journal of Computer Assisted Learning*, *41*(3), e70056. Doi: 10.1111/jcal.70056