JLER
Journal of Language Education Research

Volume 9 Number 1, January 2026

E – ISSN 2621-0169

A COMPARATIVE STUDY OF THE UTILIZATION OF RETRIEVAL-AUGMENTED GENERATION (RAG) AND CONVENTIONAL GENERATIVE AI MODELS IN IMPROVING STUDENTS' SCIENTIFIC WRITING SKILLS

Aurelia Sakti Yani[1], Restu Bias Primandhika[2]

[1]IKIP Siliwangi, Cimahi
[1]saktiyani@ikipsiliwangi.ac.id, [2]restu@ikipsiliwangi.ac.id

ABSTRACT

The rapid advancement of generative artificial intelligence (GenAI) has reshaped the higher education landscape, raising concerns about the integrity of academic writing, including issues such as plagiarism, hallucination, and the decline of students' critical thinking skills, primarily associated with conventional generative models or Large Language Models (LLMs). This paper aims to conduct a comparative analysis between the use of LLMs and the Retrieval-Augmented Generation (RAG) framework, with a case study focusing on the NotebookLM platform, specifically to identify RAG's potential in enhancing academic integrity and facilitating authentic sourcing. Employing a quantitative approach, the research targeted data collection from 37 student respondents from the Indonesian Language Education Study Program at IKIP Siliwangi through a structured questionnaire. The hypothesis posits that the RAG framework can effectively integrate information from external knowledge bases to address three main issues: 1) the way students obtain references in scientific writing, 2) the reduction of source hallucination in writing, and 3) the encouragement of students to become active knowledge curators.

Kata Kunci: Retrieval-Augmented Generation (RAG), Generative AI, NotebookLM, Scientific Writing Skills, Academic Integrity, Critical Thinking, Education

ABSTRAK

Kemajuan pesat kecerdasan buatan generatif (GenAI) telah membentuk kembali lanskap pendidikan tinggi, menimbulkan kekhawatiran tentang integritas penulisan akademik, termasuk isu-isu seperti plagiarisme, halusinasi, dan penurunan keterampilan berpikir kritis siswa, terutama terkait dengan model generatif konvensional atau Model Bahasa Besar (LLM). Makalah ini bertujuan untuk melakukan analisis komparatif antara penggunaan LLM dan kerangka kerja Retrieval-Augmented Generation (RAG), dengan studi kasus yang berfokus pada platform NotebookLM, khususnya untuk mengidentifikasi potensi RUG dalam meningkatkan integritas akademik dan memfasilitasi sumber otentik. Dengan menggunakan pendekatan kuantitatif, penelitian ini menargetkan pendataan dari 37 responden mahasiswa dari Program Studi Pendidikan Bahasa Indonesia IKIP Siliwangi melalui kuesioner terstruktur. Hipotesis ini mengemukakan bahwa kerangka kerja RAG dapat secara efektif mengintegrasikan informasi dari basis pengetahuan eksternal untuk mengatasi tiga masalah utama: 1) cara siswa memperoleh referensi dalam penulisan ilmiah, 2) pengurangan halusinasi sumber dalam tulisan, dan 3) dorongan siswa untuk menjadi kurator pengetahuan yang aktif.

***Kata kunci**: Retrieval-Augmented Generation (RAG), AI Generatif, NotebookLM, Keterampilan Menulis Ilmiah, Integritas Akademik, Berpikir Kritis, Pendidikan*

JLER
Journal of Language Education Research
Volume 9 Number 1, January 2026
E – ISSN 2621-0169

INTRODUCTION

The rapid development of generative artificial intelligence (GenAI) has brought about significant transformations in the Indonesian higher education landscape. The presence of this technology, particularly large language models (LLMs) such as ChatGPT, has transformed the way students access information, construct arguments, and complete academic assignments (Ministry of Education, Culture, Research, and Technology, 2024; Supriyadi, 2024). While GenAI offers significant potential for increasing efficiency and productivity, its presence has also raised serious concerns regarding academic integrity, particularly in the context of student academic writing (LLDIKTI 3, 2024).

One of the main challenges faced by Indonesian students is the low quality of academic writing and their weak ability to manage credible library sources (Syazali et al., 2025; Khaerunnisa, 2020; Yanti et al., 2024). Research shows that students' scientific writing skills are still relatively low, especially in developing ideas and searching for references that are relevant to the topic (Syazali et al., 2025; Baruza, 2025). This condition is exacerbated by the rampant practice of plagiarism among students, which not only threatens academic integrity but also demonstrates students' inability to produce original work (Ministry of Defense of the Republic of Indonesia, 2024). The 2024 Internal Control System (SPI) Education Survey revealed that 43% of respondents stated that plagiarism still occurs on campus, indicating that this problem has become a systemic issue in Indonesian higher education (Corruption Eradication Commission, 2025).

The use of conventional LLM in an academic context raises three crucial, interrelated issues. First, students tend to rely directly on AI output without verifying the sources underlying the information (Abbas, 2023). Second, the phenomenon of AI hallucinations—where models generate information that appears convincing but is in fact incorrect, fictitious, or even fabricates citations—poses a serious threat to the credibility of scientific work (IBM, 2023). Third, excessive reliance on AI is feared to undermine students' critical thinking skills, which should be a core competency in academia. Critical thinking skills are crucial for students because they enable them to analyze, evaluate, and synthesize information in depth (Semarang State University, 2022; Muhammadiyah

University of Jakarta, 2024), given that these skills still need to be improved in Indonesia (Sadeli, 2023; Pelita Harapan University, 2023).

To overcome the limitations of conventional LLM, an innovative approach called Retrieval-Augmented Generation (RAG) emerged. RAG integrates the capabilities of large language models with mechanisms for real-time information retrieval from external knowledge bases (InterSystems, 2025). In this way, RAG enables AI models to generate more accurate, up-to-date, and verifiable responses by linking generative outputs with authoritative data sources. Unlike conventional LLM, which relies solely on parametric knowledge from training data, RAG can access up-to-date information from multiple sources (InterSystems, 2025).

One promising implementation of RAG in the educational context is Google's NotebookLM platform, specifically designed to support research and learning activities in higher education (Google Support, 2025). NotebookLM allows students to upload various academic sources and generate summaries, study guides, and even quizzes based on those sources, thus encouraging them to become active curators of knowledge (Detik, 2025; IKIP Siliwangi, 2025).

Previous studies have shown that the implementation of RAG in various educational domains has yielded positive results. Fanani (2025) developed a RAG-based final project title evaluation system that achieved a high level of accuracy, while Hidayat et al. (2025) implemented a RAG chatbot for a student information system that was capable of providing informative and accurate responses. These studies indicate that RAG can improve the quality of AI output while reducing the risk of hallucinations through an externally sourced verification mechanism.

However, research specifically comparing the effectiveness of RAG with conventional LLM in the context of improving students' scientific writing skills, particularly in Indonesia, is still very limited. Most previous studies have focused more on the general use of AI in academic writing (Syahputra et al., 2025) or the technical implementation of RAG in areas other than writing skill development. Therefore, this study seeks to fill this literature gap by conducting a comparative analysis between the use of conventional LLM and the RAG framework, specifically through a case study of the NotebookLM platform, in the context of improving students' scientific writing skills.

This study aims to identify the potential of RAG in addressing three main problems faced by students: (1) how students obtain and manage valid and credible library sources in scientific writing; (2) reducing the phenomenon of source hallucinations; and (3) encouraging students to become active curators of knowledge with strong information literacy skills (Rahmawati, 2021). By focusing on students of the Indonesian Language Education Study Program at IKIP Siliwangi as research subjects, this study is expected to provide an empirical contribution to the understanding of how RAG technology can be integrated ethically and responsibly in the Indonesian higher education ecosystem, while maintaining and improving students' academic integrity and critical thinking skills.

METHOD

This study uses a quantitative approach with a quasi-experimental one-group pretest-posttest design . According to Sugiyono (2018), a quantitative approach focuses on numerical and statistical analysis to test formulated hypotheses. This pretest-posttest design was chosen to measure changes in students' abilities before and after being given treatment using different artificial intelligence (AI) platforms, namely conventional LLM and Retrieval-Augmented Generation (RAG). Creswell (2014) stated that this design is suitable for systematically measuring the effect of treatment variables on dependent variables.

The population in this study was all students of the Indonesian Language Education Study Program at IKIP Siliwangi. The sample was determined using purposive sampling, with the criteria being students currently taking research methodology courses or writing scientific papers, to ensure the research results are more relevant and focused. Sugiyono (2018) stated that purposive sampling was chosen based on certain considerations so that the data obtained could answer the research objectives specifically. The main instruments in this study consist of:

1. Scientific Writing Measurement

Using a scientific paper assessment rubric adapted from Syazali et al. (2025) and Khaerunnisa (2020). The rubric assesses aspects of reference integrity, scientific argument strength, and compliance with scientific publication standards on a scale of 1–4. Assessments were conducted on student work from the pretest , posttest using conventional LLM, and posttest using RAG.

2. Non-Test Instrument (Perception Questionnaire)

A 1–5 Likert scale questionnaire was designed to measure students' perceptions regarding the ease of obtaining scientific references, the accuracy of sources, and their perceptions of their role as knowledge curators (Sugiyono, 2018). The validity and reliability of the questionnaire were tested first using item analysis techniques and a minimum Cronbach's alpha of 0.70.

3. Data Collection Procedures

Data collection was carried out in four stages:

a. Pretest (Week 1): students write a scientific outline without AI assistance, then the work is assessed using a rubric as baseline data.
b. Conventional LLM Use (Weeks 2–3): students are given training in using ChatGPT for source searching and writing, then compile scientific papers and fill out perception questionnaires.
c. Use of RAG/NotebookLM (Weeks 4–5): students are given training in using NotebookLM, including uploading scientific sources and writing scientific papers with the features provided.
d. Interviews and Evaluation (Week 6): semi-structured interviews were conducted with 10 selected students to explore their experiences.

Quantitative data were analyzed using SPSS version 26. The analysis was carried out through the following steps:

a. Descriptive statistics to describe changes in student scientific scores and perceptions.
b. Normality test using Shapiro-Wilk for each data set (Sugiyono, 2018).
c. Test homogeneity with Levene's test to ensure equality of variance.
d. Paired sample t-test to compare scores before and after treatment, with a significance criterion of $\alpha = 0.05$ (Creswell, 2014).

JLER
Journal of Language Education Research
Volume 9 Number 1, January 2026
E – ISSN 2621-0169

RESULT AND DISCUSSION

Result

Before presenting the results, it should be emphasized that this study aims to comparatively test the effectiveness of conventional LLM and the RAG (NotebookLM) framework in improving the scholarly quality of students' writing and their perceptions of the knowledge curation process. Measurements were conducted using a scholarly writing rubric and a perception questionnaire, in accordance with the methodology described previously.

Table 1. Average Score of Scientific Writing Rubric (N = 37)

Dimensions	Pretest (M, SD)	Posttest LLM (M, SD)	Posttest RAG (M, SD)
Reference Integrity	2.11 (0.45)	2.78 (0.52)	3.42 (0.38)
The Power of Scientific Arguments	2.05 (0.48)	2.64 (0.55)	3.35 (0.40)
Publication Rules Compliance	2.18 (0.47)	2.71 (0.50)	3.29 (0.42)
Total Score (12 points)	6.34 (1.20)	8.13 (1.38)	10.06 (1.05)

Paired sample t-tests showed a significant increase from the LLM Pretest to the Posttest, $t(36) = 8.21$, $p < .001$, and from the RAG Pretest to the Posttest, $t(36) = 14.37$, $p < .001$. The difference between the RAG Posttest and the LLM Posttest was also significant, $t(36) = 6.54$, $p < .001$.

Table 2. Average Perception Questionnaire Score (1–5)

Variables	LLM Posttest (M, SD)	Posttest RAG (M, SD)
Ease of Getting References	3.21 (0.67)	4.02 (0.48)
AI Source Accuracy Trust	2.89 (0.72)	4.11 (0.44)
Perception of the Role of Knowledge Curator	3.05 (0.65)	4.18 (0.50)

Discussion

The results of the study showed that the use of conventional LLM (such as ChatGPT, Gemini, or Perplexity) improved all aspects of the scientific writing, but more significant improvements occurred after students used RAG (NotebookLM or LLM with upload feature). Scores for reference integrity, strength of scientific arguments, and adherence to publication rules increased consistently, indicating that the external retrieval mechanism in RAG is able to provide more authoritative sources and assist in citation validation (Fanani, 2025; InterSystems, 2025). These findings are in line with the studies of Syazali et al. (2025) and Khaerunnisa (2020) which emphasize the importance of relevant and valid library sources in writing scientific papers.

Improvements in the quality of student writing after using RAG indicate that integrating external knowledge bases into the writing process can strengthen information literacy and critical thinking skills (Baruza, 2025; Rahmawati, 2021). Students are not only consumers of information but also active in the knowledge curation process, in accordance with the demands of 21st-century learning (Semarang State University, 2022). Studies by Hidayat et al. (2025) and Syazali et al. (2025) also confirm that research-based learning and the use of AI technology integrated with external sources can improve the quality of argumentation and the structure of students' scientific writing.

Questionnaire data shows that students found it easier to obtain references and had greater confidence in the accuracy of sources after using RAG compared to conventional LLM. However, when using conventional LLM, there is still a tendency to copy and paste without verifying sources, as well as a tendency to accept information without critical processing, as revealed by Supriyadi (2024) and OpenAI (2025). This phenomenon reinforces the findings of the Corruption Eradication Commission (2025) and the Ministry of Defense (2024) regarding the prevalence of plagiarism and low academic integrity among Indonesian students.

Conversely, the use of RAG encourages students to be more active in selecting, verifying, and integrating library sources, thus enhancing knowledge curation behavior (Baruza, 2025; Fanani, 2025). This supports information literacy

and fosters a more responsible academic character, as suggested by Rahmawati (2021) and Pelita Harapan University (2023).

The results of this study confirm that integrating RAG into scientific writing learning can be an effective solution for improving writing quality, strengthening academic integrity, and fostering active knowledge curation behavior. This study also recommends continued instructional support so that students do not rely solely on AI but also develop critical thinking and information literacy skills independently (Syazali et al., 2025; Rahmawati, 2021; Baruza, 2025).

CONCLUSION

This study confirms that the Retrieval-Augmented Generation (RAG) model through NotebookLM is significantly more effective than conventional LLM in improving the quality of scientific writing of students of the Indonesian Language Education Study Program at IKIP Siliwangi. The results of measurements using the scientific rubric showed an increase in the average score from 6.34 (pretest, without AI), to 8.13 (LLM posttest), and a peak of 10.06 after using RAG, in the aspects of reference integrity, argument strength, and adherence to scientific publications. Statistically, the difference between the LLM posttest and the RAG posttest was significant ($p < 0.001$), proving the real contribution of external knowledge base integration in helping source verification and validation. Perception questionnaire data showed that students felt it was easier to obtain references (score 4.02/5) and had more confidence in the accuracy of sources (4.11/5) after using RAG, compared to conventional LLM (3.21 and 2.89, respectively). This study also noted a behavioral shift: from a tendency to accept AI output without verification in conventional LLMs, to an active role as knowledge curators when using RAG. The implications of these results are important for the higher education ecosystem: RAG integration is recommended as a solution to sustainably build information literacy, academic integrity, and critical thinking in students. Further studies are needed to test the effectiveness of this model across disciplines and analyze instructional support strategies so that AI technology truly encourages independent thinking, rather than simply technical dependence.

REFERENCES

Abbas, A. (2023). Survey analysis of the use of Artificial Intelligence (AI) in writing final scientific papers (TA-KTI) at the Ternate Computer Science Academy (AIKOM) Campus, North Maluku, Indonesia. Student Proceedings , 1 (1). https://doi.org/10.56983/prosidingkemahasiswaan.v1i1.1461

Baruza, BI (2025). Student information literacy in searching for thesis references at UIN Ar-Raniry Banda Aceh. UIN Ar-Raniry Repository . http://repository.ar-raniry.ac.id/id/eprint/41930/

Fanani, I. (2025). Implementation of Augmented Generation Retrieval for evaluating student final project proposals. Journal of Computer Technology and Informatics , 3 (2). https://jurnal.politeknikpajajaran.ac.id/index.php/tekomin/article/download/336/127/1133

Hidayat, LR, et al. (2025). Chatbot Retrieval-Augmented-Generation based on Large Language Model for optimizing student information system services. Atma Journal of Information and Communication Technology , 9 (1). https://jtika.if.unram.ac.id/index.php/JTIKA/article/view/459

IBM. (2023). What are AI hallucinations? https://www.ibm.com/id-id/think/topics/ai-hallucinations

InterSystems. (2025). Retrieval Augmented Generation (RAG): What is it and how does RAG prevent AI errors? https://www.intersystems.com/id/resources/retrieval-augmented-generation-rag-apa-itu-dan-bagaimana-rag-mencegah-kesalahan-ai/

Ministry of Education, Culture, Research, and Technology. (2024). Guidelines for the use of Generative Artificial Intelligence in learning at universities . https://lldikti3.kemdikbud.go.id/wp-content/uploads/2024/11/Buku-Panduan-_-Penggunaan-Generative-AI-pada-Pembelajaran-di-Perguruan-Tinggi-cetak.pdf

Ministry of Defense of the Republic of Indonesia. (2024). The impact of plagiarism on higher education in Indonesia. https://www.kemhan.go.id/balitbang/2024/09/17/dampak-plagiarisme-terhadap-pendidikan-tinggi-di-indonesia.html

Khaerunnisa, K. (2020). Improving scientific writing skills through a collaborative approach. Stilistika: Journal of Language, Literature, and Its Teaching , 5 (1). https://journal.um-surabaya.ac.id/index.php/Stilistika/article/view/3762

Corruption Eradication Commission. (2025). Findings from the 2024 Education SPI: Cheating and plagiarism are still rampant in schools and campuses. https://www.kpk.go.id/id/ruang-informasi/berita/temuan-hasil-spi-pendidikan-2024-menyontek-dan-plagiarisme-masih-merebak-di-sekolah-dan-kampus

LLDIKTI 3. (2024). ADIA-Integritas-Akademik. https://lldikti3.kemdikbud.go.id/adia-integritas-akademik/

OpenAI. (2025). Why language models hallucinate. https://openai.com/id-ID/index/why-language-models-hallucinate/

Rahmawati, NA (2021). Improving students' information literacy skills through information literacy classes. Islamic University of Indonesia Library Bulletin . https://journal.uii.ac.id/Buletin-Perpustakaan/article/download/20239/11614/55467

Sadeli, L. (2023). Measuring students' critical thinking skills through case study learning. Journal of the National Seminar on Continuing Education and Science , 6 (1). https://journal.unuha.ac.id/index.php/spbs/article/download/2806/793/6674

Supriyadi, E. (2024). Using OpenAI's ChatGPT in Student Scientific Writing. SDGs Proceedings . https://unars.ac.id/ojs/index.php/prosidingSDGs/article/view/4799

Syahputra, F., et al. (2025). Evaluation of the effectiveness of generative AI in helping teachers develop learning materials in Indonesia. Journal of Community Service and Educational Research , 3 (3), 265–272. https://doi.org/10.31004/jerkin.v3i3.381

Syazali, M., et al. (2025). Assessment of students' scientific writing skills through the implementation of a research-based learning model assisted by example media, templates, and paper writing guidelines. Journal of Educational Sciences and Learning , 9 (3). https://jipp.unram.ac.id/index.php/jipp/article/download/3147/1680/16565

Google Support. (2025). Getting Started with Gemini and NotebookLM for Education. https://support.google.com/a/answer/16350447?hl=en

Detik. (2025). Google launches NotebookLM Plus, its use for learning and teaching. https://www.detik.com/edu/detikpedia/d-7713335/google-luncurkan-notebooklm-plus-ini-gunanya-untuk-belajar-dan-mengajar

IKIP Siliwangi. (2025). NotebookLM - A more scientific and personal AI, a new friend for your thesis. https://it.ikipsiliwangi.ac.id/2025/03/19/notebooklm-ai-yang-lebih-ilmiah-dan-personal-teman-baru-untuk-skripsian/

Muhammadiyah University of Jakarta. (2024). Students must think critically, is that true? https://umj.ac.id/opini/mahasiswa-harus-berpikir-kritis-apakah-benar/

Semarang State University. (2022). The importance of improving critical thinking skills of 21st-century students. https://fip.unesa.ac.id/pentingnya-meningkatkan- mampu-berpikir-kritis-mahasiswa-abad-21/

Pelita Harapan University. (2023). Critical thinking skills of students in Indonesia: Low or high? K-PIN Bulletin , 9 (02). https://buletin.k-pin.org/index.php/arsip-artikel/1200-cakap-berpikir-kritis-siswa-di-indonesia-rendah-atau-tinggi

Indonesian Wikipedia. (2024). Hallucination (artificial intelligence). https://id.wikipedia.org/wiki/Hallucination_(artificial_intelligence)

Yanti, et al. (2024). Research-based learning to improve scientific writing skills . University of Mataram.