The Application of Artificial Intelligence in an Automatic Scientific Article Summarization System with a Gemini AI Engine

Gde Brahupadhya Subiksa[1],
I Nyoman Eddy Indrayana[2], I Putu Astya Prayudha[3],
I Ketut Gede Sudiartha[4], and Yessi Aprilia Waluyo[5]

[1,2,3,4,5] Technology Information Department, Politeknik Negeri Bali, Bali, Indonesia
brahupadhya@pnb.ac.id

Abstract. Students at Politeknik Negeri Bali often face difficulties when completing their final projects, especially in reviewing large numbers of scientific articles. The process of searching, reading, and understanding relevant literature is time-consuming and may hinder learning efficiency and research quality. To address this, an AI-based automatic summarization system was developed to help students quickly grasp the core content of academic articles, thereby facilitating a more efficient understanding of the literature. However, existing AI models, including ChatGPT, still have limitations in summary accuracy, contextual understanding, recognition of domain-specific terminology, and adaptability to diverse article types, often resulting in incomplete or fragmented summaries. This study aims to: (1) develop an accurate, comprehensive, and adaptive AI-based academic article summarization system; (2) design the system to meet the needs of Politeknik Negeri Bali students in final project completion; and (3) evaluate its effectiveness in improving thesis writing efficiency and quality. Effectiveness was assessed using the System Usability Scale (SUS) with 10 statements rated on a 5-point Likert scale. Fifty-four valid responses were obtained from sixth-semester Software Engineering Technology students. The system achieved an average SUS score of 80.1, categorized as "Excellent" and interpreted as "Outstanding" in user experience, earning an A grade (Very Good). Non-functional testing indicated responsive performance and fast processing for most requests. Overall, the system demonstrated excellent usability and is considered a valuable academic tool for supporting students in understanding and efficiently processing scientific literature.

Keywords: Artificial Intelligence (AI), Automatic Summarization System, Scientific Article, System Usability Scale (SUS)

1 Introduction

In today's digital era, scientific information is growing rapidly, with thousands of new articles published each year. Students face significant challenges when completing their final projects due to the time and effort required to filter and understand relevant

A. A. N. G. Sapteka et al. (eds.), *Proceedings of the International Conference on Sustainable Green Tourism Applied Science - Engineering Applied Science 2025 (ICOSTAS-EAS 2025)*, Advances in Engineering Research 280,
https://doi.org/10.2991/978-94-6463-878-3_13

content, which could potentially lower the quality of research and learning efficiency. Automatic summarization systems offer a promising solution. By leveraging Artificial Intelligence (AI), these systems can efficiently analyze and summarize scientific articles, enabling students to access essential information and quickly enhance their comprehension of complex material (Ningrum et al., 2025).

Politeknik Negeri Bali, committed to educational quality, must adopt modern technologies to support learning (Pratiwi et al., 2024; Ambara et al., 2024). Implementing AI-based summarization systems is expected to help students access relevant, high-quality resources for their final projects and improve academic performance.

However, current AI summarization models, such as ChatGPT, still face limitations. Summary accuracy is often affected by linguistic complexity and scientific nuance. These models struggle with deep contextual understanding and paragraph-level coherence, leading to fragmented or incomplete summaries (Mawlidy et al., 2024). Their grasp of domain-specific terminology and complex concepts remains limited, and adaptability to diverse article types (e.g., reviews, empirical studies, theoretical papers) is suboptimal, often requiring model retraining, which demands substantial data and computing resources (Pamungkas et al., 2024).

This research aims to address those challenges by developing a more advanced and adaptive summarization system, improving accuracy, context awareness, and handling of scientific terminology. Usability evaluation, involving direct user testing, will assess how easy and efficient the system is for end-users (Pramono et al., 2019; Suria, 2024). The outcome is expected to make a significant contribution to educational technology at Politeknik Negeri Bali and provide a practical solution for navigating scientific information overload.

2 Methodology

2.1 Related Research

Research conducted by Evan Averill Andika (2024) on the development of Artificial Intelligence (AI)-based systems in the healthcare sector found that one of the key benefits of AI is the increased efficiency of healthcare operations. AI has influenced various aspects of medical services, ranging from diagnosis to patient care, including its application in disease detection (Andika & Renaldy, 2024). In line with this, previous research by Ondra Eka Putra on the implementation of AI in hospital monitoring of patients revealed that AI can assist in developing systems that facilitate emergency nurse calls and patient condition monitoring (Putra, 2020).

The use of AI has also extended into the education sector. Hia (2024) discussed the design of an AI-based system for scheduling lectures, emphasizing that AI enables logical and systematic processes to create faster, easier, and more optimal scheduling solutions (Hia et al., 2024). Similarly, in earlier research, Selvina O. Thun (2023) investigated the development of AI-based learning materials related to the respiratory system in living organisms. The study found that AI-based teaching materials were of

satisfactory quality and suitable for use both inside and outside the classroom. Students responded positively, and AI was shown to assist learners in better understanding the subject matter (Selvina et al., 2023).

Theoretical studies on AI in higher education have also been conducted, including research by Sehan Rifky (2024) on the impact of Artificial Intelligence in higher education. The study emphasized that AI will have a significant influence on the future of higher education. AI enables personalized learning, where systems can tailor learning experiences to meet the individual needs of each student, thereby enhancing learning effectiveness. Overall, the study concluded that AI positively impacts higher education by enhancing personalized learning, enabling automated assessments, and increasing management efficiency in educational processes (Rifky, 2024).

Grace Yulianti's study titled "Transforming Indonesian Education: Harnessing the Potential of Artificial Intelligence (AI)" highlighted the role of AI in curriculum development, personalized learning, teacher-student interaction, and addressing educational inequality. The study suggests that AI can enhance teaching and learning activities by offering innovative solutions to improve learning outcomes and bridge educational gaps. AI presents a promising path toward creating a more inclusive and equitable education system in Indonesia (Yulianti et al., 2023).

Based on the findings of the aforementioned studies, it can be preliminarily concluded that AI has the potential to address various challenges in today's modern era, including those faced by Politeknik Negeri Bali, particularly in supporting students in completing their final projects. The novelty of this research lies in the development of a new web-based AI system that provides a more accessible, practical, and efficient solution for students, enabling them to save time on reading and understanding scientific articles.

2.2 Artificial Intelligence (AI)

The concept of Artificial Intelligence (AI) has existed since the mid-20th century, when scientists began exploring the possibility of creating machines capable of thinking and learning. In 1956, the Dartmouth Conference marked the formal beginning of AI research, during which the term "artificial intelligence" was first introduced. Since then, advancements in technology and algorithms have enabled significant progress in the field, including the development of sophisticated models that can understand and process human language. With improvements in computing power and the availability of big data, AI has rapidly evolved, opening new opportunities for innovation and application across various domains (Jaya et al., 2018).

Artificial Intelligence (AI) is a branch of computer science focused on developing systems and technologies capable of performing tasks that typically require human intelligence. AI encompasses various techniques and methods enabling machines to learn, adapt, and make decisions based on available data. With these capabilities, AI can solve complex problems while enhancing efficiency and productivity across various sectors, including healthcare and industry. As technology advances, AI has become an integral part of daily life, influencing how we interact with the world around us (Kaira, 2022).

One valuable tool in the AI ecosystem is the OpenAI Python library, designed to facilitate developers in accessing and utilizing artificial intelligence models developed by OpenAI, including language models like GPT (Generative Pre-trained Transformer) (Assegaf et al., 2024). By utilizing this library, users can seamlessly integrate AI capabilities into their applications, including natural language processing, text generation, and dialogue-based interactions. The library offers a straightforward and intuitive interface, facilitating rapid and efficient API calls. Additionally, the OpenAI Python library supports various features, including model parameter settings, session management, and result processing, allowing users to customize their AI experience to meet specific needs. With comprehensive documentation and an active community, this library is a popular choice among developers exploring AI potential in both commercial and research projects. Its usage also reflects a broader trend in software development, where AI integration is becoming increasingly essential for creating more innovative and more responsive solutions (Budiman et al., 2023).

2.3 Scientific Article

A scientific article is a written work based on systematic and structured research or scholarly review. Its primary purpose is to communicate research findings, analyses, or new ideas to the academic and scientific community (Judijanto et al., 2024). Such articles are typically written by researchers, academics, or professionals in a specific field and published in accredited scientific journals. The standard structure of a scientific article includes essential sections such as introduction, literature review, methodology, results, and discussion, all contributing to a comprehensive understanding of the topic addressed (Widiyastuti et al., 2023).

Scientific articles can be categorized into several types based on their purpose, methods, and content. First, research articles are the most common, where authors report original research results, including methodology, data analysis, and findings. Second, review articles provide a comprehensive analysis of existing studies in a field, aiming to summarize recent developments and identify research gaps. Third, methodology articles focus on developing or testing new research methods, offering guidance on techniques for other researchers. Fourth, case study articles provide an in-depth analysis of a specific case, often used in health or business fields to illustrate best practices or challenges. Finally, opinion or editorial articles express the author's views or arguments on current issues within a field, often based on their experience or expertise (Udil, 2021). Each type of scientific article plays a vital role in enriching academic literature and promoting scholarly discussion (Darmalaksana & Hambali, 2021).

The importance of scientific articles lies not only in delivering new information but also in their contribution to the advancement of science and technology. Articles published in accredited journals undergo a peer-review process, where experts in the field evaluate the quality and validity of the research before publication. This process ensures that only high-quality research is accepted, making scientific articles reliable references for other researchers and academics (Sari et al., 2023). Thus, scientific

articles play a crucial role in expanding knowledge, fostering scholarly discussion, and facilitating progress across various disciplines.

2.4 Research Stages

This section explains the stages of the research to be conducted. The stages of this research, as shown in Figure 1, can be described as follows.

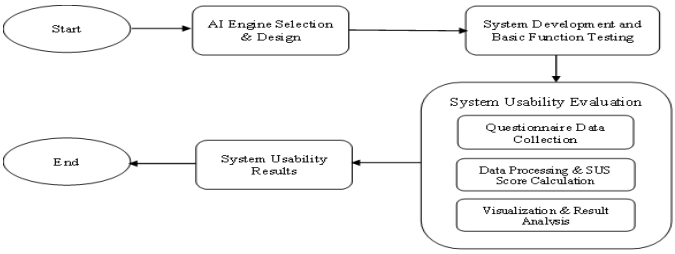

Figure 1. Research Stages Explains the Stages of Research

3 Result and Discussion

3.1 Result

Comparison of AI engines. The selection of the AI engine to be used requires several considerations related to specific criteria (Maulida, 2024). Currently, there are three AI engines in use, namely ChatGPT, Gemini AI, and DeepSeek. A comparison of these three engines can be seen in Table 1 as follows:

Table 1. Comparison of the Three AI Engines Planned for Use

Criteria	ChatGPT (OpenAI)	Gemini AI (Google)	DeepSeek AI
Support for Indonesian Language	Good	Very Good (native multilingual)	Fairly Good
Accuracy of Scientific Summaries	High (with GPT-4)	High + more stable long-context handling	Fair (more focused on coding)
Number of Free API Hits	0 (requires paid API subscription)	Generous (1000–1500 requests/month)	Limited
API Response Time	Fast	Fast + supported by Google	Slow in some trials
Input Token Limit	~8k (GPT-3.5) / ~32k (GPT-4)	1 million characters (via Gemini 1.5)	Max 16k tokens
Python Integration	Easy (using OpenAI library)	Easy (using google.generativeai)	Not yet stable

Based on the technical and functional analysis in Table 1, Gemini AI from Google is chosen as the main engine for this system because it offers advantages in Indonesian language support, a large free quota, stable performance, and the capacity to process long documents. This engine is highly suitable to support an automatic summarization system accessed by students, lecturers, and researchers on a wide scale.

Development and Testing of the System's Core Functions. This stage explains how the system is developed and how each of its main functions is tested to ensure they work properly according to user needs. The system was successfully built with a simple and user-friendly interface (Figure 2), providing ease of use from the user's perspective.

Figure 2. The System was Successfully Built with A Simple and User-Friendly Interface (in Indonesian

Basic System Function Testing (Blackbox). Basic system function testing is conducted using the Blackbox method, which is a software testing approach that focuses on inputs and outputs without examining the internal structure of the program code (Jibril, 2024). The purpose of this testing is to ensure that the main features of the AI-Based Automatic Scientific Article Summary system function according to the specifications and user requirements (Table 2).

Table 2. Basic System Function Testing (Blackbox)

Feature Tested	Input	Expected Process	Expected Output	Test Result	Remarks
Upload a PDF File	Valid scientific article PDF file	The system reads the file, extracts text, and displays metadata and article content	Metadata displayed, such as journal name, author, year, and article content	Passed	Works as intended
File Format Validation	Non-PDF file format (.docx/.jpg)	The system rejects the file and shows an error message	Error message: "File must be in PDF format".	Passed	Validation works well
Article Metadata Extraction	A PDF containing the journal name and author	The system detects the journal name, author, and year from the PDF text	Metadata successfully displayed on the user interface	Passed	Accurate extraction

Feature Tested	Input	Expected Process	Expected Output	Test Result	Remarks
AI Summary with Gemini	Extracted article text	The system sends text to the Gemini API for automatic summarization	Structured summary: Introduction, Method, Results, and Conclusion	Passed	Summary format correct

Non-Functional Testing. Non-functional testing aims to measure system aspects beyond basic functionality, such as the ability to handle load, stability, and scalability. One important parameter tested is the token output limit, which in Gemini 2.0 Flash Free is set to a maximum of 8,192 tokens. In this test, the output of the summarization system has been adjusted to display only the background, methods, discussion, and conclusion sections, with a limitation of 1 to 5 sentences per section. Based on the use of the free Gemini 2.0 Flash API (Monitoring Google API/Service Details), the non-functional testing was conducted (Table 3).

Table 3. Non-Functional Testing

Request	Errors	Avg latency	99th percentile latency
1,116	8.15%	0.886 seconds	14.724 seconds

The results of the non-functional testing show that 1,116 requests were processed by the system, with 8.15% of the total requests experiencing errors or failures. The average time required by the system to process a request and provide a response was 0.886 seconds, indicating that the system has responsive performance and can deliver results quickly for most requests. However, the latency for 99% of the requests remains noteworthy, with 14.724 seconds being the processing time threshold for 99% of the requests, indicating that a small portion of requests required more time to be processed. Overall, these results demonstrate that the Gemini AI-based automatic summarization system operates effectively for most requests. However, there is room for optimization to reduce the error rate and improve latency for slower requests.

3.2 Discussion

In the performance testing of the automatic summarization system, 1,116 requests were processed with an error rate of 8.15% and an average latency of 0.886 seconds. While the majority of requests were processed quickly, 99% of the requests had a latency equal to or lower than 14.724 seconds, indicating that 1% of the requests experienced longer delays. This suggests that although the system's overall performance is generally good, further usability evaluation will be essential to assess the broader feasibility of using the system. To provide a more straightforward overview of the usability evaluation results using the System Usability Scale (SUS), a visualization was created in the form of a bar chart displaying the SUS scores of each respondent. This

visualization facilitates the identification of user perception patterns regarding the developed system (Figure 3).

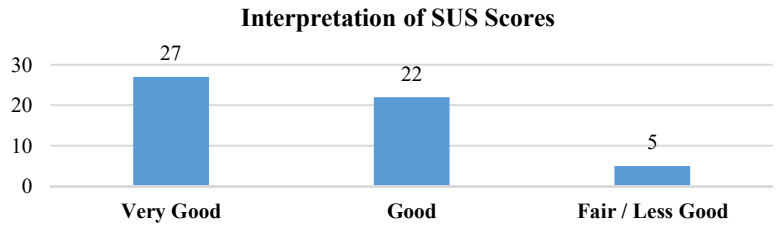

Figure 3. Visualization Chart of SUS Method Test Results Based on Respondents' Answers

The distribution of individual scores also indicates that the majority of respondents gave scores above 80, with some respondents achieving very high scores (95-97), suggesting that the user experience with this system is generally positive (Figure 4).

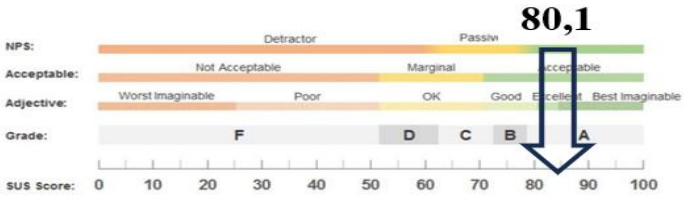

Figure 4. Interpretation of the Visualization of SUS Method Results

Based on Figures 3 and 4, the AI-based Automatic Scientific Article Summarization System exhibits an excellent usability level. It has high potential for broader adoption by students, primarily to support academic literature reviews. Some minor improvements may still be necessary to achieve optimal and maximum performance. Referring to the SUS score of 80.1, the Automatic Scientific Article Summarization System can be categorized into several usability evaluation classifications. First, the score of 80.1 on the Net Promoter Score (NPS) dimension falls within the Promoter range (80–100), indicating that users are delighted and highly likely to recommend the system to others. Second, from the Acceptability perspective, this score falls within the Acceptable category (70–100), indicating that the system is already suitable for real-world use by users. Third, in terms of subjective rating (Adjective Rating), the score falls within the Excellent level (75–85), indicating that users' experience with the system is highly satisfying. Lastly, from the letter grade perspective (Grade), the score of 80.1 is classified as Grade A (>79): Very Good. This interpretation suggests that the developed system has successfully delivered a positive user experience and is suitable for application in an academic context.

4 Conclusion

The results of the non-functional testing indicate that the system processed 1,116 requests, with 8.15% of these requests encountering errors or failures. The average time required for the system to process a request and provide a response was 0.886 seconds, indicating responsive performance and the ability to deliver results quickly for most requests. Additionally, based on the usability evaluation using the System Usability Scale (SUS) method, the AI-Based Automatic Scientific Article Summarization System achieved an average score of 80.1 from 54 respondents, who were sixth-semester students of the Software Engineering Technology Study Program (S1 TRPL). The evaluation process involved distributing a questionnaire consisting of 10 statements rated on a 1–5 Likert scale. The questionnaire data were processed and calculated using the standard SUS formula, resulting in the following findings: (a) The SUS score of 80.1 indicates that the system has an excellent level of usability. (b) The Net Promoter Score (NPS) of 80.1 falls within the Promoter range (80–100), indicating that users are delighted and likely to recommend the system to others. (c) The Acceptability Range classifies the system as Acceptable, meaning it is deemed suitable and practical for use in an academic context. (d) The Adjective Rating corresponds to Excellent, reflecting a highly satisfying user experience with the system. (e) The Grade achieved is an A, signifying that the system meets usability standards very well from the users' perspective. The SUS data visualization also shows that most users rated the system as easy to use, trustworthy in its features, and efficient in the workflow. Furthermore, the category distribution reveals a dominant perception of "Very Good" among users toward the tested system. Therefore, it can be concluded that the AI-Based Automatic Scientific Article Summarization System has successfully met user expectations in terms of usability and is suitable for use as an academic tool, especially in facilitating students', lecturers', and researchers' understanding of scientific article content.

References

Ambara, M. P., Sudiarta, M., & Suryaniadi, S. M. (2024). Digitalization test materials and information system for scheduling competency test sites at Politeknik Negeri Bali's Professional Certification Institute. *International Conference on Sustainable Green Tourism Applied Science-Social Applied Science 2024 (Icostas-Sas 2024),* 856–865.

Andika, E. A., & Renaldy, R. (2024). Perkembangan sistem berbasis artificial intelligence di bidang kesehatan. *Prosiding Seminar Nasional Informatika*, *2*, 593–598.

Assegaf, I. A., Syastra, M. T., Rifky Kurniawan, T., & Soleh, M. (2024). Pengembangan chatbot konsultasi kesehatan mental kesehatan mental berbasis OpenAI model GPT-3.5 turbo menggunakan media Whatsapp. *Jurnal Informatika Teknologi dan Sains (Jinteks), 6*(4), 785–793.

Budiman, A. W., Setiawan, A., & Nugroho, S. (2023). Pengembangan sistem layanan informasi berbasis web dengan memanfaatkan AI pada ChatGPT. *Jurnal Teknologi dan Sistem Informasi Bisnis, 5*(4), 592–602.

Darmalaksana, W., & Hambali, R. (2021). Penulisan tugas akhir skripsi dalam bentuk artikel ilmiah: Studi kasus Fakultas Ushuluddin UIN Sunan Gunung Djati Bandung. *Pre-Print Kelas Menulis UIN Sunan Gunung Djati Bandung*, 1–8.

Hia, A. S., Andika, N., & Supratikta, H. (2024). Perancangan sistem informasi penjadwalan kuliah berbasis AI (artificial intelligence). *Multidisciplinary Indonesian Center Journal (Micjo), 1*(3), 1198–1206.

Jaya, H., Sabran, S., Idris, M., Djawad, Y. A., Ilham, A., & Ahmar, A. S. (2018). *Kecerdasan buatan*. Fakultas MIPA Universitas Negeri Makassar.

Jibril, M. (2024). Pengujian sistem informasi e-modul pada SMPN 1 Tempuling menggunakan black box testing. *Jurnal Perangkat Lunak, 6*(2), 327–332.

Judijanto, L., Wibowo, G. A., Hakpantria, H., Nuryanneti, I., Apriyanto, A., Firdaus, A., Haryanti, T., Zafar, T. S., Jumiono, A., & Jati, P. (2024). *Karya tulis ilmiah: Panduan praktis menyusun karya tulis ilmiah*. PT. Sonpedia Publishing Indonesia.

Kaira, M. F. (2022). *Implementasi model bahasa OpenAI GPT-3 untuk aplikasi text content generator berbasis web dan aplikasi mobile*. Digital Repositori Unila.

Maulida, R. (2024). Komparasi respons ChatGPT dan Gemini terhadap command pattern identik dengan metode black box. *Jurnal Teknik Informatika, 10*(2), 68–71.

Mawlidy, E. R., Dio, R., & Lorensa, L. (2024). Kemampuan artificial intelligence terhadap pendeteksian fraud: Studi literatur. *Akurasi: Jurnal Studi Akuntansi dan Keuangan, 7*(1), 89–104.

Ningrum, T. S., Anspratiwi, V. D., & Wahyudi, M. (2025). Etika dan dampak sosial dari penerapan artificial intelligence dalam sistem informasi manajemen. *Jurnal Riset Sistem Informasi, 2*(1), 6–15.

Pamungkas, B. A., Dharmawan, A., & Yusup, Y. (2024). Development of Android education application for Arabic alphabet with artificial intelligence for children. *Sistemasi: Jurnal Sistem Informasi, 13*(4), 1449–1463.

Pramono, W. A., Az-Zahra, H. M., & Rokhmawati, R. I. (2019). Evaluasi usability pada aplikasi MyTelkomsel dengan menggunakan metode usability testing. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 3*(3), 2951–2959.

Pratiwi, N. P. D., Sanjaya, I. B., & Adiningrat, G. P. (2024). Analisis kecepatan dan kehandalan pelayanan jasa perawatan dan perbaikan terhadap tingkat kepuasan pelayanan unit penunjang akademik perawatan dan perbaikan Politeknik Negeri Bali. *Jurnal Syntax Admiration, 5*(9), 3512–3524.

Putra, O. E. (2020). Implementasi artificial intelligence pada sistem pengawasan pasien rumah sakit. *Jurnal Teknologi, 10*(2), 28–41.

Rifky, S. (2024). Dampak penggunaan artificial intelligence bagi pendidikan tinggi. *Indonesian Journal of Multidisciplinary on Social and Technology, 2*(1), 37–42.

Sari, V. B. M., Solichin, M. R., Sholikha, S. M., Priastuti, D. N., Cahyani, F., & Noviana, N. E. (2023). Peningkatan kualitas penulisan artikel ilmiah terakreditasi Sinta melalui pelatihan interaktif. *Community Development Journal: Jurnal Pengabdian Masyarakat, 4*(5), 10269–10275.

Selvina, O., Setyaningrum, V., & Atmaja, D. S. (2023). Pengembangan bahan ajar ChatBot berbasis artificial intelligence pada materi sistem pernapasan makhluk hidup di sekolah dasar. *Muallimuna: Jurnal Madrasah Ibtidaiyah, 8*(2), 88–99.

Suria, O. (2024). A statistical analysis of system usability scale (SUS) evaluations in online learning platform. *Journal of Information Systems and Informatics, 6*(2), 992–1007.

Udil, P. A. (2021). Pelatihan penulisan artikel ilmiah penelitian tindakan kelas untuk publikasi pada jurnal ilmiah. *Jurnal Nasional Pengabdian Masyarakat, 2*(1), 21–27.

Widiyastuti, N. E., Sanulita, H., Waty, E., Qani'ah, B., Purnama, W. W., Tawil, M. R., Hermawan, M. S., Zahara, A. E., Asry, W., & Daniati, D. (2023). *Inovasi & pengembangan karya tulis ilmiah: Panduan lengkap untuk penelitian dan mahasiswa*. PT. Sonpedia Publishing Indonesia.

Yulianti, G., Bernardi, B., Permana, N., & Wijayanti, F. A. K. W. (2023). Transformasi pendidikan Indonesia: Menerapkan potensi kecerdasan buatan (AI). *Journal of Information Systems and Management (Jisma), 2*(6), 102–106.