ORIGINAL ARTICLE

pISSN 1226-6396, eISSN 2234-4942
J of Oriental Neuropsychiatry 2025;36(4):251-263
https://doi.org/10.7231/jon.2025.36.4.251

JON Journal of Oriental Neuropsychiatry

Student Usability of an AI-Based Process-Oriented Approach for EBM Education in Korean Medicine

Chan-Young Kwon*,†, Min-Jae Kim*

*Department of Oriental Neuropsychiatry, Dong-eui University College of Korean Medicine, †Anti-Aging Research Center, Dong-eui University

Received: November 10, 2025
Revised: December 10, 2025
Accepted: December 12, 2025

Correspondence to
Chan-Young Kwon
Department of Oriental Neuropsychiatry, Dong-eui University College of Korean Medicine, 52-57 Yangjeong-ro, Busanjin-gu, Busan, Korea.
Tel: +82-51-850-8808
Fax: +82-51-867-5162
E-mail: beanalogue@naver.com

Acknowledgement
We would like to express our sincere gratitude to the fourth-year students at the Dong-eui University College of Korean Medicine who participated in this study for their time and valuable feedback.

Objectives: Evidence-based medicine (EBM) education often emphasizes the critical evaluation of published research as a final product, neglecting the dynamic and ongoing nature of scientific inquiry. Our recent proof-of-concept study highlighted the educational potential of an AI-based research note generation system from the perspective of faculty educators. However, the actual learning experience and effectiveness from the students' viewpoint remain unexamined.

Methods: This study assessed the usability and perceived educational value of an AI-based research note generation system for Korean medicine (KM) students. A cross-sectional study involved 36 fourth-year students who used the system for an assignment. The assignment required students to upload clinical research papers, review 10 AI-generated research notes, answer three AI-generated questions, and receive AI feedback. Evaluation methods included a mixed-use questionnaire and the EBM interest rating scale, which was scored on a 0∼10 scale before and after the task.

Results: Participants reported high overall satisfaction, with the highest scores in "Usefulness and Convenience" (4.53±0.43) and "System Usability" (4.51±0.51). After completing the assignment, students' interest in EBM increased significantly (from 5.25 to 7.19, $p < 0.001$). Qualitatively, students valued the "proxy experience" of engaging in the research process, appreciated the time efficiency of the system, and critically noted issues with AI content accuracy.

Conclusions: This study indicates that an AI-based, process-centered learning approach is a feasible and promising strategy for EBM education, significantly enhancing student motivation and interest. While students recognized the advantages of using this system, they also critically acknowledged the limitations of AI, particularly concerning content accuracy.

Key Words: Artificial intelligence, Evidence-based medicine, Medicine education, Educational technology, Medical students.

Copyright © 2025 by The Korean Society of Oriental Neuropsychiatry. All rights reserved.
This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. INTRODUCTION

Evidence-based medicine (EBM) integrates clinical expertise, patient values, and the best external evidence to aid medical decision-making and is considered a core paradigm in modern medicine[1]. Accordingly, EBM education has become an essential component of today's medical curricula[1]. However, the optimal strategy for teaching EBM to undergraduate medical students remains controversial, and previous studies have reported inconsistent results[2-4]. The primary approach in traditional EBM education focuses on critical evaluation of published research articles[2]. While essential for developing critical appraisal skills, this approach has limitations, as it treats research as a static end product, obscuring the dynamic and complex process by which knowledge is created[5,6]. This limitation is particularly pronounced in fields such as Korean medicine (KM), where research often involves complex, multi-component interventions or bridges traditional paradigms with modern methodologies[7]. For these students, understanding the context and process of research, not simply evaluating the final results, is crucial for navigating the complexities of integrating traditional knowledge with evidence-based practice. For learners, the intellectual journey from initial research questions and methodological debates to inevitable setbacks and discoveries often remains a "black box."

This results-oriented approach can lead to a superficial understanding of scientific inquiry, preventing learners from developing a deep and contextual understanding of the research process itself. Therefore, a pedagogical shift is needed, moving beyond the static analysis of completed papers to a more holistic, process-oriented, and longitudinal perspective[8]. This approach does not necessarily mean extending the curriculum across multiple years, but rather focuses on reconstructing the complete narrative of a single research project. By following the journey from initial question to final conclusion, learners can understand the context and complexity of scientific work, a crucial skill for navigating the ambiguity of clinical reasoning[9]. As clinical decision-making becomes increasingly complex, this need for process-oriented learning becomes increasingly important.

The recent emergence of generative artificial intelligence (AI) offers opportunities to realize these educational models. In medical education, generative AI is already being utilized to build adaptive learning platforms, virtual patient scenarios, and interactive simulations that enhance personalized learning[10]. However, the application of these powerful tools has primarily focused on clinical skills and student-instructor interactions[10]. To date, few AI tools are designed to simulate the complex cognitive and decision-making processes of researchers. If AI can systematically decompose final publications and replicate researchers' assumed thought processes, tasks, and decisions, it could offer an innovative solution to the limitations of existing EBM education[8].

Recently, our research team developed an educational system that utilizes generative AI to reconstruct published academic papers into a series of simulated research notes. A proof-of-concept study was conducted to evaluate the educational utility and potential of this system for professional educators in EBM and research methodology[8]. While the faculty involved acknowledged the conceptual value of the system, they raised concerns about the reliability of the AI content and its practical application in real-world classrooms[8]. However, the ultimate value of an educational tool is determined by the learner's actual experiences and outcomes. While faculty perspectives are important, direct evidence is needed to understand how students experience and perceive these AI-based tools.

Therefore, this study aimed to evaluate the us-

ability and perceived educational utility of this AI-based system for KM students. Specifically, we examined students' learning experiences using the system in an actual classroom setting, their usability, their perceptions of AI functions, and changes in their interest in EBM. This study aims to elucidate the practical effectiveness and limitations of a process-oriented, AI-based, evidence-based medicine education approach from the learner's perspective.

II. METHODS

1. The educational system

The web-based educational tool evaluated in this study was developed on the Google Apps Script platform and leverages the Google Gemini 2.5 Flash API to automatically generate research notes. The model was specifically chosen for its rapid response speed and cost-efficiency, which are critical for scalable educational tool designed for simultaneous classroom use. For student accessibility, the system's web interface was hosted on GitHub Pages, while the backend logic was handled by Google Apps Script (Fig. 1A).

The system architecture consists of four core components: (1) a paper analysis engine, (2) a research note generation engine, (3) a question generation engine, and (4) a teaching support module. The paper analysis engine utilizes the Google Drive API to perform optical character recognition and natural lan-

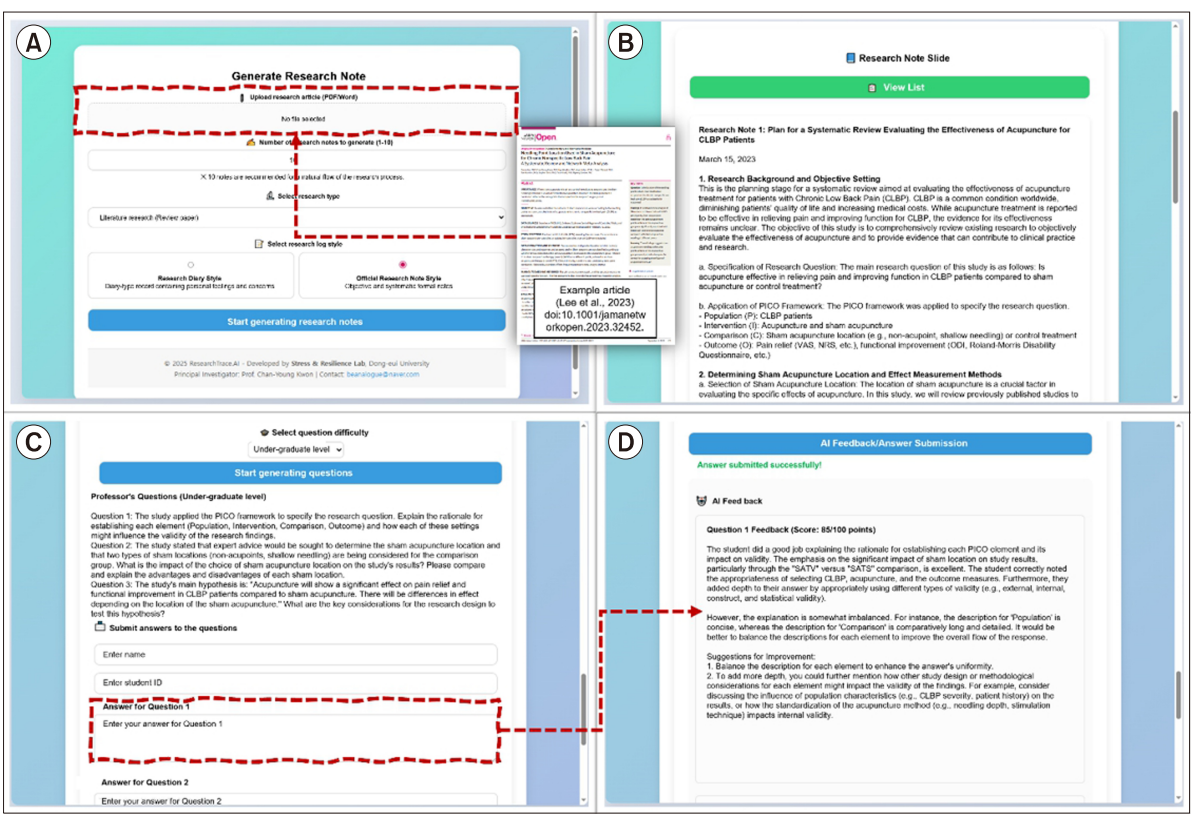


Fig. 1. User interface and workflow of the AI-based research note generation system. The figure illustrates the four main stages of the student's interaction with the system, corresponding to the assignment steps. (A) The "Generate Research Note" screen, where a student uploads a research paper and configures settings. (B) An example of a "Research Note Slide" (Note 1) generated by the AI, which simulates the research process. (C) The "Teaching Support Module" screen, displaying the three AI-generated undergraduate-level discussion questions. (D) The "AI Feedback" screen, showing the detailed, AI-generated evaluation and feedback provided after the student submits their answers.

guage processing on uploaded research papers (PDF or Word), extracting key elements such as research objectives and methodology to structure the input data. The research note generation engine applies customized algorithms to reconstruct the chronological progression of the research. To ensure logical consistency, the engine utilizes a sequential prompt engineering strategy: it first establishes the overall research timeline and then generates notes step-by-step. Specific prompt constraints were implemented to strictly simulate the decision-making stages; for example, the system is programmed to restric information about future results or conclusions in early-stage notes to provide an authentic, process-centered learning experience. The question generation engine creates discussion questions categorized into difficulty levels based on the generated notes. Finally, the teaching support module provides AI-based feedback on student responses, evaluating them against a predefined scoring rubric to support personalized learning. In this study, questions at the undergraduate level were used (Fig. 1B).

2. Study design and participants

This study employed a cross-sectional, descriptive, anonymous survey design. Data were collected via an online survey over approximately one week, from November 3 to November 10, 2025. The selection criteria were as follows: 1) fourth-year students at D University College of KM during the semester of 2025, 2) students who completed required assignments using the system, and 3) students who voluntarily agreed to participate. Students who did not complete the assignments or did not consent to participation were excluded.

3. Educational intervention and assignment structure

As part of the curriculum, students were given two weeks to complete an assignment consisting of the following steps: 1) Paper Selection: Each team selected a clinical research paper of interest. 2) System Setup and Paper Upload: Students accessed the system website, uploaded the selected paper (PDF file), set the number of research notes to 10, and selected the research type and writing style for the paper. 3) Research Note Review: Students reviewed 10 sequential research notes generated by the system and examined the decision-making process at each step. 4) Question and Answer: Students responded to three automatically generated undergraduate-level discussion questions and submitted their answers. 5) AI Feedback Review: Students reviewed AI-generated evaluations and feedback on their submitted answers. A practical guide detailing the assignment process was provided to students. They were required to enter their name and student ID number to verify their completion of the assignment (Fig. 2).

4. Survey instrument

The research team developed a new questionnaire for this study based on a literature review and expert opinions. To ensure content validity, the initial draft was reviewed by two professors of KM and one clinical research expert. Based on their feedback, ambiguous expressions were refined to ensure clarity for undergraduate students. The questionnaire consisted of three parts (Supplementary File 1).

1) Respondent Demographic Information (4 items): This section collected data on gender, age, self-reported frequency of using generative AI (e.g., ChatGPT, Claude) for learning purposes, and interest in EBM and clinical research (on a 0～10 scale) before and after the assignment.

2) Quantitative Assessment (30 items): This section assessed the following six domains using a 5-point Likert scale (1=strongly disagree, 5=strongly agree): (a) Learning Motivation and Attitude (7 items); (b)

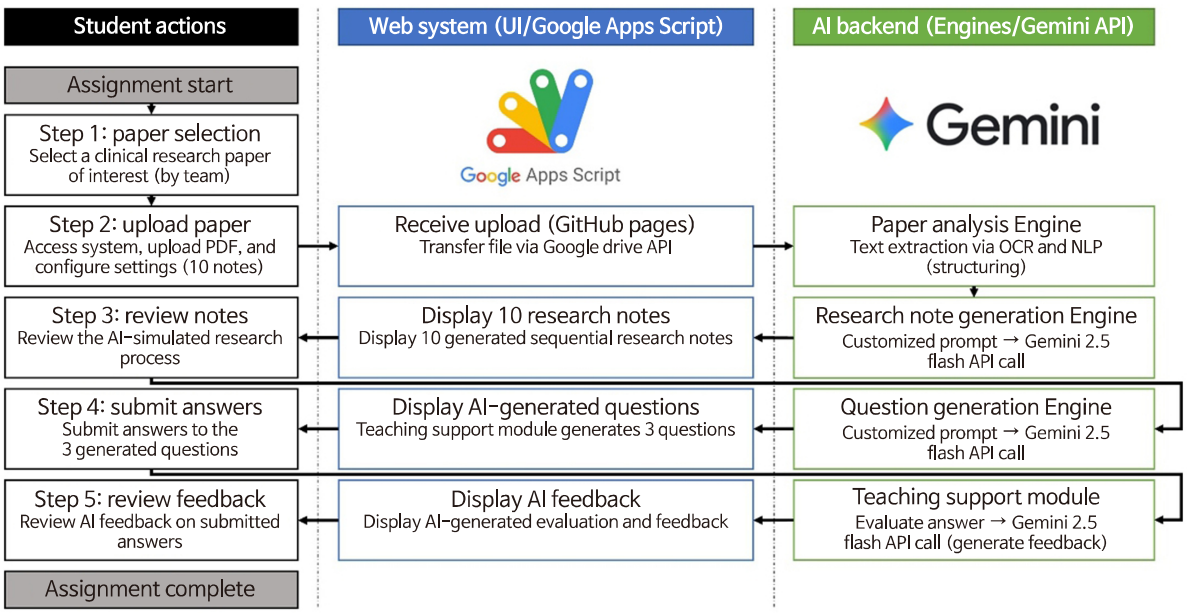


Fig. 2. Flowchart of the educational intervention and system architecture. This diagram illustrates the step-by-step procedure of the assignment, mapping student actions to the system's front-end and back-end processes. The flow is divided into three lanes: (1) Student Actions: The 5-step assignment process, from paper selection to feedback review. (2) Web System (UI/Google Apps Script): The front-end interface (hosted on GitHub Pages) that displays information and receives student input. (3) AI Backend (Engines/Gemini API): The back-end components, including the Paper Analysis Engine (OCR/NLP), Research Note Generation Engine (Gemini 2.5 Flash API call), Question Generation Engine, and the Teaching Support Module for feedback generation.

Perceived Learning Outcomes (6 items); (c) Usability and Convenience (5 items); (d) System Usability (4 items); (e) AI-Generated Questions (4 items); and (f) AI Feedback (4 items).

3) Qualitative Feedback (3 items): This section included open-ended questions regarding aspects participants were satisfied with, dissatisfied with, and their suggestions for system improvement.

5. Ethical considerations

This study was approved by the IRB of Dong-eui University Korean Medicine Hospital (Approval No: DH-2025-20, approved on October 24, 2025). All participants provided written informed consent after receiving a full explanation of the study. They were assured that participation was voluntary, would not affect their academic performance, and that they could withdraw at any time without penalty. All data was anonymized using study codes. Mobile phone numbers were collected solely for distributing a 5,000 KRW coffee e-gift and were deleted immediately afterward.

6. Data analysis

Quantitative data were analyzed using descriptive statistics. Continuous variables from Likert-scale items were expressed as means and standard deviations (SD). Categorical demographic data were summarized as frequencies and percentages. Changes in EBM interest before and after the task were analyzed using paired-samples t-tests. Qualitative data from open-ended questions were analyzed using a conventional content analysis to categorize key themes and opinions. The responses were reviewed to identify recurring keywords and grouped into broader themes to supplement the quantitative findings[11]. All statistical analyses were performed using Python 3.10 (Python Software Foundation) and Microsoft Excel 365 (Microsoft Corp., Redmond, WA, USA).

III. RESULTS

1. Respondent characteristics

All 36 students who completed the assignment participated in the survey (100% response rate). The gender distribution was 15 males (41.7%) and 21 females (58.3%). The mean age of the respondents was 25.4±2.05 years (range: 23 to 31). Regarding the frequency of using generative AI for learning, none of the respondents answered "never use," 8.3% (n=3) "rarely use (1～2 times a month)," 25.0% (n=9) "occasionally use (1～2 times a week)," 47.2% (n=17) "frequently use (3～4 times a week)," and 19.4% (n=7) "very often use (almost every day)" (Table 1).

2. Quantitative evaluation results

The quantitative evaluation results across the six domains are presented in Table 2. The domain with the highest score was "Usefulness and Convenience" (4.53±0.43), followed by "System Usability" (4.51±0.51), "AI Feedback" (4.41±0.54), "Perceived Learning Outcomes" (4.36±0.50), and "Learning Motivation and Attitude" (4.32±0.47). The 'AI-Generated Questions' domain received the lowest score (4.19±0.56), though it still indicated a high level of positive assessment.

Among individual items, the highest-rated were: '*It was convenient to use because the system is a non-face-to-face service*' (Q17: 4.67±0.48), '*I could save time analyzing papers by participating in the assignment*' (Q15: 4.61±0.64), and '*The menus and terminology used in the system were not difficult*' (Q21: 4.61±0.55). Items with relatively lower scores included: '*My interest in EBM increased after this system assignment*' (Q3: 3.94±0.67), '*The difficulty level of AI-generated questions was appropriate for "undergraduate level"*' (Q24: 3.97±0.84), and '*I trust the content of the feedback provided by AI*' (Q30: 4.14±0.76). Nonetheless, these items also received mean scores near or above 4.0.

In the "Learning Motivation and Attitude" domain, high scores were given to: '*The assignment using this system was interesting*' (Q1: 4.44±0.56), '*It is appropriate to use this system assignment in this course*' (Q5: 4.42±0.65), and '*I support using this system assignment for future students' EBM education*' (Q7: 4.39±0.55). In the "Perceived Learning Outcomes" domain, '*This system assignment helped my EBM-related knowledge (e.g., paper analysis)*' (Q8: 4.47±0.70) and '*The assignment will help me provide evidence-based care as a future KM doctor*' (Q13: 4.42±0.55) were rated particularly high. Within the "AI Feedback" domain, the item '*The AI feedback on my answers was useful*' (Q27: 4.58±0.60) received the highest score, followed by '*AI feedback helped identify parts I understood insufficiently or missed*' (Q28: 4.47±0.61) and '*AI feedback provided specific help*

Table 1. Respondent Demographic Characteristics (N=36)

Characteristic	Category	n (%)
Gender	Male	15 (41.7)
	Female	21 (58.3)
Age (years)	Mean±SD	25.4±2.05
	Range	23～31
Frequency of generative AI use for learning*	Never	0 (0)
	Rarely (1～2 times/month)	3 (8.3)
	Sometimes (1～2 times/week)	9 (25.0)
	Often (3～4 times/week)	17 (47.2)
	Very often (almost daily)	7 (19.4)

AI: artificial intelligence, SD: standard deviation.
*Refers to the use of generative AI tools such as ChatGPT and Claude for learning purposes.

Table 2. Evaluation Results by Domain

Domain	Mean±SD	Range
A. Learning motivation and attitude (7 items)	4.32±0.47	3.29～5.00
Q1. The assignment using this system was interesting	4.44±0.56	3～5
Q2. I actively participated in the assignment using this system	4.36±0.59	3～5
Q3. My interest in EBM increased after this system assignment	3.94±0.67	2～5
Q4. My interest in critically reading research papers increased after the assignment	4.33±0.63	3～5
Q5. It is appropriate to use this system assignment in this course	4.42±0.65	3～5
Q6. I recommend applying AI-based assignments like this system in other courses	4.33±0.68	3～5
Q7. I support using this systemassignment for future students' EBM education	4.39±0.55	3～5
B. Perceived learning outcomes (6 items)	4.36±0.50	3.33～5.00
Q8. This system assignment helped my EBM-related knowledge (e.g., paper analysis)	4.47±0.70	3～5
Q9. The assignment was useful for understanding my current state of EBM knowledge	4.36±0.83	2～5
Q10. The assignment helped improve my critical thinking skills	4.33±0.68	3～5
Q11. I gained confidence in identifying key contents of research papers through the assignment	4.22±0.68	3～5
Q12. The assignment will help my actual research performance ability	4.33±0.68	3～5
Q13. The assignment will help me provide evidence-based care as a future Korean medicine doctor	4.42±0.55	3～5
C. Usefulness and convenience (5 items)	4.53±0.43	3.60～5.00
Q14. This system was overall useful	4.36±0.64	3～5
Q15. I could save time analyzing papers by participating in this system assignment	4.61±0.64	3～5
Q16. I could participate in this system assignment regardless of time and place	4.47±0.65	3～5
Q17. It was convenient to use because this system is a non-face-to-face service	4.67±0.48	4～5
Q18. The guidance (practice handout) for using this system was appropriate	4.56±0.65	3～5
D. System usability (4 items)	4.51±0.51	3.00～5.00
Q19. The design (interface) of this system facilitated task performance	4.44±0.69	2～5
Q20. Accessing this system was not difficult	4.56±0.65	3～5
Q21. The menus and terminology used in this system were not difficult	4.61±0.55	3～5
Q22. I could use the system without prior knowledge of this system	4.42±0.73	2～5
E. AI-generated questions (4 items)	4.19±0.56	2.50～5.00
Q23. The three questions generated by AI reflected the key contents of my selected paper well	4.22±0.64	3～5
Q24. The difficulty level of AI-generated questions was appropriate for "undergraduate level"	3.97±0.84	1～5
Q25. AI-generated questions stimulated my thinking and made me think more deeply about the paper	4.22±0.64	3～5
Q26. AI-generated questions were clear and easy to understand	4.33±0.76	2～5
F. AI feedback (4 items)	4.41±0.54	3.00～5.00
Q27. The AI feedback on my answers was useful	4.58±0.60	3～5
Q28. AI feedback helped identify parts I understood insufficiently or missed	4.47±0.61	3～5
Q29. AI feedback provided specific help in improving my answers	4.44±0.61	3～5
Q30. I trust the content of the feedback provided by AI	4.14±0.76	3～5

AI: artificial intelligence, EBM: evidence-based medicine, SD: standard deviation.
All items were rated on a 5-point Likert scale (1=strongly disagree, 5=strongly agree). Domain scores represent the mean of individual item scores within each domain.

in improving my answers' (Q29: 4.44±0.61), which were also highly rated.

3. Changes in EBM interest

Interest in EBM and clinical research, rated on a 0～10 scale, increased from 5.25±2.41 (range: 0 to 10) before the assignment to 7.19±1.88 (range: 3 to 10) afterward. The mean change was 1.94 points. A paired t-test confirmed that this difference was statistically significant ($t=-8.377$, $p<0.001$, Cohen's $d=1.42$) (Fig. 3).

4. Qualitative analysis results

Responses to the open-ended questions were provided by 14 students (satisfied aspects), 7 students (dissatisfied aspects), and 6 students (areas for improvement). Thematic content analysis revealed three main themes.

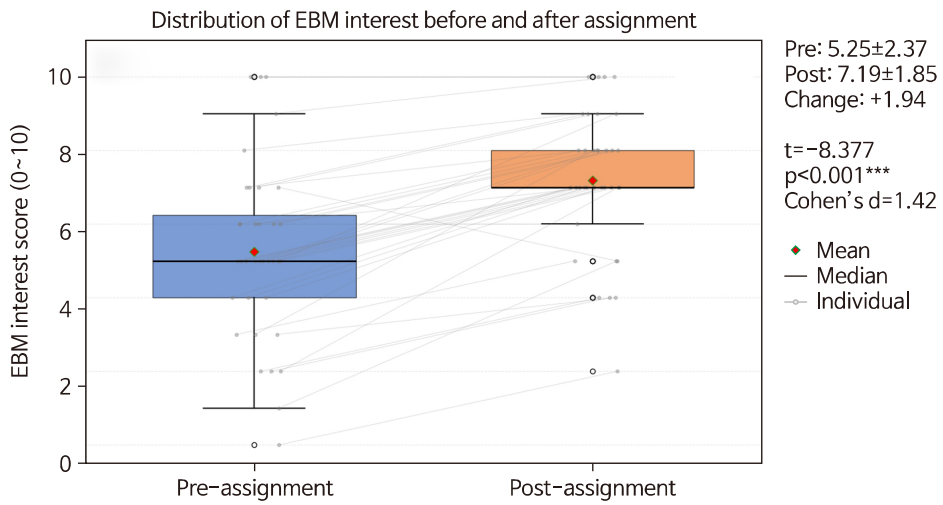


Fig. 3. Distribution of evidence-based medicine interest before and after the AI assignment. EBM interest was measured on a 0~10 scale before and after completing the AI assignment. Box plots show the median (black line), mean (red diamond), interquartile range (box), and range (whiskers). Gray lines connect individual students' scores. The change was statistically significant with a large effect size (paired t-test: t=-8.377, p<0.001, Cohen's d=1.42). ***indicates p<0.001.
EBM: evidence-based medicine.

1) Positive learning experience and practical value

Students highlighted the system's practical value and the positive learning experience. One student commented, '*It is excellent that AI is being actively used for paper analysis.*' Another noted, '*I didn't expect much because it was AI, but I was surprised that it asked sharp questions, almost like a human did! It seemed like a great assignment for many students to participate in within a limited time.*' Satisfaction with time efficiency was also high. One student remarked, '*The best part was being able to analyze a long paper in a short amount of time.*' Another said, '*It was an English paper, so I worried if the AI could summarize it well, but my concerns were unfounded. It provided perfect materials, and I could contemplate the answers based on that.*' The value of indirectly experiencing the research process was also emphasized. One response noted, '*It was an opportunity to indirectly experience the researcher's thought process during paper writing.*' Another student added, '*I had never written a paper, so the process felt vague and difficult. It was good to learn the process, and this proxy experience seems helpful for when I eventually write a paper.*' The AI feedback feature also received positive evaluations. One student opined, '*The process where the AI formed questions, graded and analyzed the answers, and then provided feedback was very novel. I think this could reduce the spatio-temporal limitations that come from interacting with a human (instructor) when students engage in research activities in the future.*' (Supplementary File 2)

2) Initial comprehension difficulty and content reliability issues

Students raised concerns regarding their initial understanding of the assignment and the reliability of the system-generated content. One student noted, '*At first, it was difficult to understand exactly what the assignment was just by reading the guide. However, once I used it, I could easily grasp the AI's purpose.*' Another said, '*It was hard to get a feel for the assignment at the beginning.*' Concerns about the accuracy

of the AI-generated content were also expressed. One pointed out, '*When the AI summarized the paper, there were content errors, such as fabricating information not present in the uploaded paper.*' Technical difficulties were also mentioned: '*Choosing the paper type at the beginning was difficult. I was confused about which category this paper belonged to.*' Difficulties stemming from a lack of background knowledge were also reported. '*I usually lack background knowledge in the research field, so it was a bit difficult to understand the assignment and the AI-generated questions right away,*' one student commented (Supplementary File 2).

3) Suggestions for system improvement

Students provided specific suggestions for functional improvements and enhanced educational support. The most prominent demand was for improved content accuracy. '*When running the AI a few times, it seemed like there were occasional errors where the keyword changed. I don't remember exactly, but for example, it felt like an error where "depression" was mentioned when the main topic was "stress".*' one student noted. Another suggested, '*A plan is needed to reduce hallucinations when the AI creates the paper summary.*' There were also suggestions for usability improvements, such as a request: '*It would be nice to have a function like an "auto-save" for answers!*' Demands for enhanced educational support were also raised. Comments included: '*It would be good if the handout clearly stated that the AI helps in "following the researcher's thought process".*' '*I would like more explanation about what the assignment is at the beginning,*' and '*It would be helpful to have examples when writing answers to the questions. It was a bit difficult for an undergraduate to critically evaluate a paper.*' (Supplementary File 2)

IV. DISCUSSION

This is the first study to evaluate the practical usability and educational potential and student perceptions of an AI-based research note generation system among KM students. Responses were collected from all 36 students who completed an assignment using the system in an actual classroom setting, providing valuable data on the practical effects of a process-oriented, AI-based EBM educational approach from the learner's perspective. Theoretically, this system aligns with the Cognitive Apprenticeship model[12], which emphasizes 'making thinking visible.' By externalizing the typically implicit decision-making processes of researchers, the system functions as a tool for modeling. Furthermore, the sequential generation of research notes acts as technological scaffolding which emphasizes 'making thinking visible.' By externalizing the typically implicit decision-making processes of researchers, the system functions as a tool for modeling. Furthermore, the sequential generation of research notes acts as technological scaffolding[13], supporting students in navigating complex inquiry tasks that would be difficult to manage independently.

The most notable finding in this study was the consistently high positive ratings across all evaluation domains (mean scores ranging from 4.19 to 4.53). This is significantly higher than the scores from faculty evaluation studies of the same systems (scores ranging from 3.70 to 3.97)[8]. This discrepancy may reflect fundamental differences in how educators and learners evaluate educational tools. Educators are often more sensitive to content accuracy, real-world classroom integration, and pedagogical relevance, whereas students tend to focus more on direct learning experiences, ease of use, and perceived effectiveness[14-16]. Notably, the "Usefulness and Convenience" domain (4.53), which corresponds to the

"Practical Applicability" domain, which received the lowest scores in faculty studies, received the highest scores in this study[8]. This suggests that students perceived the system's practical value in real-world usage contexts.

A statistically significant increase ($p<0.001$) in EBM interest before (mean 5.25) and after (mean 7.19) the assignment suggests the positive educational impact of the system. The mean increase of 1.94 points represents a 19.4% improvement on a 10-point scale, suggesting that just one task can positively change students' attitudes toward EBM. This result is particularly significant in the context of previous studies[2-4] that highlight the limitations of existing EBM instructional methods. This empirically confirms that the process-centered learning approach is not only conceptually sound but also an effective instructional strategy that can truly enhance students' motivation and interest. However, a noteworthy point regarding EBM interest is the discrepancy between the significant increase observed on the 0~10 scale and the relatively low score for item Q3, 'My interest in EBM increased after this system assignment' (3.94), which was one of the lowest-rated items. This suggests the two measures may capture different constructs. While the 0~10 scale measured a general change in 'interest' in the topic, the 5-point Likert item (Q3) asked for a stricter attribution, questioning whether this specific assignment 'caused' the increase, potentially leading students to respond more conservatively. This also aligns with the study's limitation (a single-group design), as it cannot be ruled out that the reported increase in interest may be a temporary 'novelty effect' from using a new AI tool rather than a direct result of the pedagogical approach itself.

Qualitative analysis revealed that students found value in several aspects of the system, particularly the vicarious experience of the research process. Student expressions such as '*indirectly experiencing the researcher's thought process*,' '*learning the paper writing process*,' and '*proxy experience*' suggest that the system successfully achieved its intended educational goal of enabling students to experience the longitudinal research process[5,6]. This is a learning experience that is difficult to provide with existing approaches that focus solely on analyzing completed papers.

High satisfaction with time efficiency and accessibility underpins the core advantages of AI-based educational tools. Student comments such as being '*able to analyze in a short amount of time*,' '*allowing many students to participate in a limited time*,' and the potential to 'reduce the spatiotemporal limitations' demonstrate the potential for AI to enhance educational scalability and accessibility[10]. One student's emphasis on the '*spatiotemporal limitations that come from interacting with a human (instructor)*' demonstrates the practical value that AI-based tools can provide even within realistic constraints such as limited teaching resources and increasing student numbers. Of course, as this study lacks a control group, it is difficult to definitively separate whether the high student satisfaction and increased interest stemmed from the AI-based, process-oriented approach itself or from a 'novelty effect' associated with introducing a new AI technology into the curriculum. Nevertheless, the fact that students' qualitative feedback highlighted specific experiences aligned with the core goals of process-oriented learning, such as the '*proxy experience of the researcher's thought process*' and '*learning the paper writing process*', suggests the system likely delivered its intended educational value beyond mere novelty.

However, this study also confirmed that some of the concerns raised in the faculty study[8] were reflected in the student experience. Specifically, students recognized issues with the accuracy of AI-gen-

erated content, such as hallucinations. Some critiques, such as '*content errors, such as fabricating information not present in the uploaded paper*' and '*errors where the keyword changed*,' align with the content reliability concerns faculty raised[8]. Interestingly, the relatively low score for '*I trust the content of the feedback provided by AI*' (Q30: 4.14) suggests that students are aware of the AI's limitations and are approaching it critically. This can be interpreted positively, as it indicates students are not blindly trusting the AI tool but are maintaining their critical evaluation skills. This critical stance is likely correlated with high levels of AI literacy. As the results show, two-thirds of students (66.6%) reported using generative AI "often" or "very often" for learning. These students are experienced users who already understand the inherent limitations of AI, such as hallucinations, and therefore do not blindly accept AI output. This implies that the participants in this study may have 'early adopter' characteristics, already being familiar with AI technology. Consequently, their high acceptance and critical awareness might not be representative of other student populations less familiar with AI. Therefore, further validation is needed to determine if these positive findings can be generalized to students with less prior experience in using generative AI.

It is also worth noting the relatively low score (3.97) for the item '*The difficulty level of AI-generated questions was appropriate for "undergraduate level"*' (Q24). This may suggest a mismatch between the students' actual proficiency and the AI's predefined "undergraduate" difficulty setting. Qualitative responses such as '*It was a bit difficult for an undergraduate to critically evaluate a paper*' or '*it was a bit difficult to understand... the AI-generated questions right away*' suggest the need for difficulty adjustments due to a lack of background knowledge. This suggests that more detailed customization beyond simple undergraduate/master's/doctoral classifications may be necessary to accommodate students' diverse backgrounds and experience levels.

Feedback indicating initial difficulty understanding the task highlights the importance of instructional design. Comments such as '*it was difficult to understand exactly what the assignment was' and 'it was hard to get a feel for the assignment*' demonstrate that innovative educational tools require clear instructions and explanations. This aligns with the need for '*clear learning objectives and a structured implementation guide*' identified in the faculty study[8]. Student suggestions such as '*clearly stating... that the AI helps in "following the researcher's thought process"*' and '*providing examples for answers,*' offer specific directions for future improvement in instructional design.

Integrating these findings with previous faculty evaluations[8] confirms the validity and feasibility of this process-oriented, AI-based EBM educational approach. Faculty acknowledged its conceptual value while raising concerns about reliability and practicality[8]. Conversely, students reported high satisfaction and learning outcomes in real-world applications. This dual validation validates the system's educational potential while clearly identifying specific areas for improvement. Within the broader context of the discussion on the role of AI tools in medical education[17,18], this study demonstrates that AI can function beyond mere information delivery or automation. This suggests that AI can serve as an educational partner that fosters students' higher-order thinking and process understanding. This is particularly important in fields such as KM, which often involve complex, multicomponent interventions and individualized treatments[8], where understanding the complexity and context of research becomes even more crucial. The system evaluated here contributes by presenting this complexity to students in an ac-

cessible manner.

Despite these results, this study has several important limitations. First, the generalizability of the results is limited because the participants were from a single institution and a single grade level (fourth year). Furthermore, it is important to acknowledge the potential 'novelty effect.' Since 66.6% of the participants were already frequent users of generative AI, the positive responses may partially reflect their enthusiasm for new technology rather than the pedagogical approach itself. These 'early adopter' characteristics might limit the generalizability of the findings to students who are less familiar with or more resistant to AI tools. Second, this cross-sectional study only assessed short-term perceptions and satisfaction and did not assess long-term learning outcomes or actual research skills. Third, the reliance on self-report questionnaires raises the possibility of social desirability bias. Fourth, because the assignment was part of a graded, regular course, students' evaluations may not fully reflect their voluntary perspectives. Fifth, the study employed a single-group design without a control group, making direct comparisons with traditional paper-reading methods difficult. Sixth, the measurement of EBM interest relied on a single self-report item, which did not objectively measure actual improvements in EBM knowledge or skills. Finally, given the rapid advancement of AI technology, the results using a specific AI model (Gemini 2.5 Flash) may not apply to future, more advanced models.

Future research should expand in several directions. First, studies across multiple institutions, grade levels, and medical disciplines are needed to verify generalizability. Second, longitudinal designs are needed to assess long-term learning outcomes and actual research capabilities. Third, randomized controlled trials should be conducted to directly compare the system's effectiveness with existing teaching methods. Fourth, objective assessment tools should be developed to measure real-world changes in knowledge, skills, and attitudes. Fifth, technical improvements are needed to increase the accuracy of AI content and minimize hallucinations. Sixth, personalized difficulty adjustment algorithms based on students' background knowledge should be developed. Seventh, implementation guidelines should be developed, integrated with faculty training programs, to improve practical classroom integration. Finally, comparative studies using more advanced AI models (e.g., GPT-4, Claude 3) are needed to evaluate the educational improvements resulting from technological advancements.

V. CONCLUSION

This study examined the practical usability and perceived educational utility of an AI-based process-oriented learning system from the perspective of KM students. Participants reported high satisfaction with the system's usability and convenience, particularly with its statistically significant increase in interest in EBM. Students highly valued the system's ability to provide an indirect "*proxy experience*" of the research process, which is difficult to achieve with conventional methods. However, they also critically recognized limitations, including the inaccuracy of the AI content. While this AI-based process-oriented approach appears to be a practically applicable and promising strategy for increasing students' motivation and interest in EBM, continued technical (e.g., accuracy) and pedagogical (e.g., clear instructions) improvements are needed.

CONFLICT OF INTERESTS

None.

FUNDING

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-Innovative Human Resource Development for Local Intellectualization program grant funded by the Korea government (MSIT) (IITP-2025-RS-2020-II201791).

SUPPLEMENTARY MATERIALS

Supplementary materials can be found via https://doi.org/10.7231/jon.2025.36.4.251

REFERENCES

1. Nielsen LD, Løwe MM, Mansilla F, Jørgensen RB, Ramachandran A, Noe BB, et al. Interventions, methods and outcome measures used in teaching evidence-based practice to healthcare students: an overview of systematic reviews. BMC Med Educ. 2024;24(1):306. 10.1186/s12909-024-05259-8
2. Bello JO, Grant P. A systematic review of the effectiveness of journal clubs in undergraduate medicine. Can Med Educ J. 2023;14(4):35-46. 10.36834/cmej.72758
3. Ahmadi SF, Baradaran HR, Ahmadi E. Effectiveness of teaching evidence-based medicine to undergraduate medical students: a BEME systematic review. Med Teach. 2015;37(1):21-30. 10.3109/0142159x.2014.971724
4. Tikkinen KAO, Guyatt GH. Understanding of research results, evidence summaries and their applicability-not critical appraisal-are core skills of medical curriculum. BMJ Evid Based Med. 2021;26(5):231-3. 10.1136/bmjebm-2020-111542
5. Mathieson K, Weemer M, Lipke L. Approaches to teaching evidence-based medicine in residency: a systematic review. Med Educ Online. 2025;30(1):2504467. 10.1080/10872981.2025.2504467
6. Simons M, Rapport F, Zurynski Y, Stoodley M, Cullis J, Davidson AS. Links between evidence-based medicine and shared decision-making in courses for doctors in training: a scoping review. BMJ Open. 2022;12(4):e057335. 10.1136/bmjopen-2021-057335
7. Yin CS, Ko SG. Introduction to the history and current status of evidence-based korean medicine: a unique integrated system of allopathic and holistic medicine. Evid Based Complement Alternat Med. 2014;2014:740515. 10.1155/2014/740515
8. Kwon CY. Exploring the Educational Potential of a Prototype AI System for Automated Research Note Generation: A Faculty Survey. J of Oriental Neuropsychiatry. 2025; 36(3):107-19.
9. Norman G. Research in clinical reasoning: past history and current trends. Med Educ. 2005;39(4):418-27. 10.1111/j.1365-2929.2005.02127.x
10. Tian S, Jin Q, Yeganova L, Lai PT, Zhu Q, Chen X, et al. Opportunities and challenges for ChatGPT and large language models in biomedicine and health. Brief Bioinform. 2023;25(1):1-21. 10.1093/bib/bbad493
11. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res. 2005;15(9):1277-88. 10.1177/1049732305276687
12. Collins A, Brown JS, Newman SE. Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. Knowing, learning, and instruction: Essays in honor of Robert Glaser. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1989. p. 453-94.
13. Lajoie SP. Extending the Scaffolding Metaphor. Instructional Science. 2005;33(5):541-57. 10.1007/s11251-005-1279-2
14. Kim KJ, Kang Y, Kim G. The gap between medical faculty's perceptions and use of e-learning resources. Med Educ Online. 2017;22(1):1338504. 10.1080/10872981.2017.1338504
15. Kim KJ. Medical student needs for e-learning: perspectives of the generation Z. Korean J Med Educ. 2024; 36(4):389-99. 10.3946/kjme.2024.312
16. Springmier K. An Exploration of Faculty and Student Perceptions of Generative AI. Library Trends. 2025;73(4): 426-42. 10.1353/lib.2025.a968490
17. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. Med Teach. 2024; 46(4):446-70. 10.1080/0142159x.2024.2314198
18. Saroha S. Artificial Intelligence in Medical Education: Promise, Pitfalls, and Practical Pathways. Adv Med Educ Pract. 2025;16:1039-46. 10.2147/amep.S523255

Supplementary File 1. Questionnaire

A survey on the usability and educational effectiveness of an AI-based research note generation system.

Part 1: Respondent Demographic Information

This section collects demographic data.

1. **Gender:**
    - ( ) Male
    - ( ) Female
    - ( ) Other / Prefer not to say
2. **Age:** ____ years
3. **Frequency of using generative AI (e.g., ChatGPT, Claude) for learning purposes:**
    - ( ) Never
    - ( ) Rarely (1-2 times/month)
    - ( ) Sometimes (1-2 times/week)
    - ( ) Often (3-4 times/week)
    - ( ) Very often (almost daily)
4. **Interest in EBM and clinical research:**
    - Please rate your interest on a scale of 0 (Not interested at all) to 10 (Very interested).
    - **4a. Before** completing the assignment: ____
    - **4b. After** completing the assignment: ____

Part 2: Quantitative Assessment

This section assesses your experience with the system using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree).

(A) Learning Motivation and Attitude

- **Q1.** The assignment using this system was interesting.
- **Q2.** I actively participated in the assignment using this system.
- **Q3.** My interest in EBM increased after this system assignment.
- **Q4.** My interest in critically reading research papers increased after the assignment.
- **Q5.** It is appropriate to use this system assignment in this course.
- **Q6.** I recommend applying AI-based assignments like this system in other courses.
- **Q7.** I support using this system assignment for future students' EBM education.

(B) Perceived Learning Outcomes

- **Q8.** This system assignment helped my EBM-related knowledge (e.g., paper analysis).
- **Q9.** The assignment was useful for understanding my current state of EBM knowledge.
- **Q10.** The assignment helped improve my critical thinking skills.
- **Q11.** I gained confidence in identifying key contents of research papers through the assignment.
- **Q12.** The assignment will help my actual research performance ability.
- **Q13.** The assignment will help me provide evidence-based care as a future Korean medicine doctor.

(C) Usefulness and Convenience

- **Q14.** This system was overall useful.
- **Q15.** I could save time analyzing papers by participating in this system assignment.

- **Q16.** I could participate in this system assignment regardless of time and place.
- **Q17.** It was convenient to use because this system is a non-face-to-face service.
- **Q18.** The guidance (practice handout) for using this system was appropriate.

(D) System Usability

- **Q19.** The design (interface) of this system facilitated task performance.
- **Q20.** Accessing this system was not difficult.
- **Q21.** The menus and terminology used in this system were not difficult.
- **Q22.** I could use the system without prior knowledge of this system.

(E) AI-Generated Questions

- **Q23.** The three questions generated by AI reflected the key contents of my selected paper well.
- **Q24.** The difficulty level of AI-generated questions was appropriate for "undergraduate level".
- **Q25.** AI-generated questions stimulated my thinking and made me think more deeply about the paper.
- **Q26.** AI-generated questions were clear and easy to understand.

(F) AI Feedback

- **Q27.** The AI feedback on my answers was useful.
- **Q28.** AI feedback helped identify parts I understood insufficiently or missed.
- **Q29.** AI feedback provided specific help in improving my answers.
- **Q30.** I trust the content of the feedback provided by AI.

Part 3: Qualitative Feedback

This section includes open-ended questions about your experience.

1. **Satisfied Aspects:** What aspects of using this system were you satisfied with?
2. **Dissatisfied Aspects:** What aspects of using this system were you dissatisfied with?
3. **Suggestions for Improvement:** What suggestions do you have for improving this system?

Supplementary File 2: Qualitative Feedback

Anonymous raw data from open-ended questions.

---

1. Satisfied Aspects (N=14)

ID	Response (English)
S1	It is excellent that AI is being actively used for paper analysis.
S2	I didn't expect much because it was AI, but I was surprised that it asked sharp questions, almost like a human did! It seemed like a great assignment for many students to participate in within a limited time.
S3	Although I had looked up several papers while doing assignments so far, this was the first time I analyzed a paper in such detail, so it was meaningful and I learned a lot.
S4	It was interesting that it analyzed the paper's research design process and methods and summarized the main content. It was good that it analyzed points I couldn't grasp on my own while reading the paper.
S5	The process where the AI formed questions, graded and analyzed the answers, and then provided feedback was very novel. I think this could reduce the spatiotemporal limitations that come from interacting with a human (instructor) when students engage in research activities in the future.
S6	The best part was being able to analyze a long paper in a short amount of time.
S7	It was an English paper, so I worried if the AI could summarize it well, but my concerns were unfounded. It provided perfect materials, and I could contemplate the answers based on that.
S8	I was satisfied that I could receive feedback.
S9	It was not difficult to follow.
S10	It was an opportunity to indirectly experience the researcher's thought process during paper writing.

S11	I had never written a paper, so the process felt vague and difficult. It was good to learn the process, and this proxy experience seems helpful for when I eventually write a paper.
S12	It was beneficial to learn about the process a researcher goes through.
S13	It was novel.
S14	I was able to analyze the paper effectively.

---

2. Dissatisfied Aspects (N=7)

ID	Response (English)
D1	While the AI's scoring results explained well what was lacking, it would be good if the grading criteria used to provide that feedback were disclosed.
D2	At first, it was difficult to understand exactly what the assignment was just by reading the guide. However, once I used it, I could easily grasp the AI's purpose.
D3	When the AI summarized the paper, there were content errors, such as fabricating information not present in the uploaded paper.
D4	It was hard to get a feel for the assignment at the beginning.
D5	I usually lack background knowledge in the research field, so it was a bit difficult to understand the assignment and the AI-generated questions right away.
D6	Choosing the paper type at the beginning was difficult. I was confused about which category this paper belonged to.
D7	I have always found papers difficult, but this system gave me time to organize them systematically and think about the questions.

---

3. Suggestions for Improvement (N=6)

ID	Response (English)
I1	When running the AI a few times, it seemed like there were occasional errors where the keyword

	changed. I don't remember exactly, but for example, it felt like an error where "depression" was mentioned when the main topic was "stress".
I2	It would be good if the handout clearly stated that the AI helps in "following the researcher's thought process".
I3	A plan is needed to reduce hallucinations when the AI creates the paper summary.
I4	It would be nice to have a function like an "auto-save" for answers!
I5	I would like more explanation about what the assignment is at the beginning.
I6	It would be helpful to have examples when writing answers to the questions. It was a bit difficult for an undergraduate to critically evaluate a paper.