Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

The Illusion of Mastery in AI-Assisted Learning: Do AI-Generated Summaries Create False Confidence?

Tushar Joshi
Dept. of Computer Science & Engineering
Tula's Institute of Technology
Dehradun, Uttarakhand

Dr. Bharti Kalra
Associate Professor
Dept. of Computer Science & Engineering
Tula's Institute of Technology
Dehradun, Uttarakhand

Gagan Pratap Singh
Dept. of Computer Science & Engineering
Tula's Institute of Technology
Dehradun, Uttarakhand

Shree Bhandari
Dept. of Computer Science & Engineering
Tula's Institute of Technology
Dehradun, Uttarakhand

Karan Singh Danu
Dept. of Computer Science & Engineering
Tula's Institute of Technology
Dehradun, Uttarakhand

Abstract - This paper studies how AI-generated summery can produced a false sense of understanding. Today many students use AI tools to quickly understand books, textbook, notes, report, and lecture material because the students receive an accurate and easy to understand summery that are way faster to understand then traditional learning pattern. However, reading a polished summery does not, means that real learning happens. Student may feel confident about a topic but actually unable to explain, memorise and applied it properly.

This paper examines a phenomenon called the Illusion of Mastery, which is a cognitive state where students believe they fully understand a topic after reading AI-generated summaries without deeply engaging with the original material. The study connects this problem with ideas from cognitive psychology such as metacognition, retrieval practice, and desirable difficulties, which suggest that deep learning happens through active thinking and mental effort rather than passive reading alone.

To get more accurate results, we also performed a small case study using our AI-Knowledge Workspace which works on a Retrieval-Augmented Generation (RAG) model. In this study, students used the workspace to learn technical topics through AI-generated responses and cited source material. Their performance and confidence level were later compared with students who studied using traditional textbook learning methods. The results showed that many students using AI assistance felt more confident, but their actual understanding in analytical and procedural tasks was comparatively weak.

Keywords—AI-assisted learning, illusion of mastery, false confidence, RAG.

I. INTRODUCTION

Artificial intelligence writing and summarisation tools have become a common part of everyday learning very quickly. Today many students use chatbots to prepare for exams by getting short and easy summaries instead of reading full chapters. Professionals also use AI-generated summaries to understand long reports in less time, while researchers use them to get a basic idea about unfamiliar topics before reading actual research papers. In all these situations, people often feel that they understand the topic after reading the AI summary. However, feeling like you understand something and actually understanding it properly are two very different things.

Cognitive psychology has always shown that feeling like you know something and actually being able to remember, explain, or apply it are two different things. Many studies in metacognition and memory research have discussed this gap, especially the idea called the “feeling of knowing” introduced by Hart [1]. This becomes important when students try to judge how well they are prepared for exams or academic tasks. AI-generated summaries may increase this gap even more because students often feel they understand the topic after reading a smooth and easy summary, while their actual understanding may still remain weak.

Because AI-generated summaries are usually written in a very clear, organised, and easy-to-understand way, students often develop a strong feeling that they fully understand the topic after reading them. However, this understanding may remain very shallow. The learner does not actively struggle with the material, connect new ideas with previous knowledge, solve confusion on their own, or build a strong mental understanding of the topic. They simply read the summary. Many studies in memory research have shown that only reading information is much weaker for long-term learning compared to activities like self-explanation, active recall, elaboration, and retrieval practice.

This paper studies this problem by connecting it with existing research from cognitive psychology and education. We are not trying to say that AI summaries are completely harmful, but we argue that students often ignore the learning risks that can happen when they depend too much on AI-generated content. In many cases, these summaries can create false confidence, where learners feel prepared without actually developing deep understanding of the topic.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

To examine this issue more practically, we also performed a live case study using our AI Knowledge Workspace based on a Retrieval-Augmented Generation (RAG) pipeline. The system was tested with student participants to observe how AI-assisted learning affects confidence level, understanding, and actual academic performance.

II. THEORETICAL BACKGROUND

A. Metacognition and Calibration

Metacognition—thinking about one's own thinking—consists of two broad capacities: metacognitive knowledge (beliefs about how cognition works) and metacognitive monitoring (real-time judgements about one's own understanding). Accurate monitoring, often operationalised as calibration between predicted and actual test performance, is a reliable predictor of academic achievement [2].

Calibration errors mainly happen in two ways. The first is under confidence, where students think they know less than they actually do, although this is less common and usually not very harmful. The second is overconfidence, where learners believe they understand a topic much better than they actually do. This often causes students to stop studying too early, spend less effort on revision and practice, and enter important exams with a false sense of preparation and confidence [3]. Because of this, understanding the factors that create overconfidence becomes very important in real educational situations.

Many research studies have shown that when information is easy to read and understand, people often become more confident about their learning. If the material is well-organised, polished, and written in a smooth way, students usually feel that they have understood it properly compared to content that is difficult or confusing to read [4]. AI-generated summaries work in a very similar way because they present information in a short, clean, and highly understandable format. Because of this, AI summaries may become a strong reason behind student overconfidence and false understanding during learning.

B. The Testing Effect and Generative Learning

The testing effect, also called retrieval practice, means that students learn and remember information better when they try to recall it from memory instead of only reading the same material repeatedly [5]. Many studies in memory research have shown that active recall strengthens long-term learning more effectively than simple re-reading. When students successfully remember information, their memory becomes stronger, while unsuccessful recall helps them identify weak areas that require more study and practice.

Generative learning is also connected to this idea. It includes learning activities where students actively create something from the material, such as writing summaries, giving explanations, making predictions, or connecting concepts with examples instead of only reading passively. Fiorella and Mayer [6] discussed different generative learning strategies and found that these methods usually produce better understanding and retention than traditional re-reading methods. The main reason is that students actively think about and construct meaning from the information rather than simply receiving it.

AI summarisation changes the normal learning process. In traditional learning, students usually create their own summaries after reading study material, which helps them think deeply and remember concepts better. However, with AI tools, the summary is already generated for the learner. The student mainly reads and reviews the content instead of actively constructing understanding on their own. Because generative learning depends heavily on mental effort and active participation, depending too much on AI-generated summaries may reduce the actual learning benefit even if the summary itself is accurate, clear, and well-organised.

C. Desirable Difficulties

Bjork's concept of desirable difficulties [7] explains that learning methods which feel difficult in the beginning often produce stronger understanding and long-term memory later. Techniques like spaced repetition, retrieval practice, interleaving, and self-generation are considered effective because they force students to put more mental effort into learning. Even though these methods may feel slow or uncomfortable, they usually help learners engage more deeply with the study material.

AI-generated summaries work in the opposite way because they are designed to make learning easier and faster. They simplify information, organise important points, and remove much of the difficulty from the learning process. Students no longer need to identify key ideas on their own, solve confusion, or deeply analyse the author's argument. However, these difficult mental activities are actually an important part of deep understanding and meaningful learning.

III. THE ILLUSION OF MASTERY: A MECHANISM

Through our research and study, we found a pattern that has four stages through which AI summaries may generate the illusion of mastery in learners.

Firstly, students try to understand a topic, but when the difficulty increases, they try to find an easier solution, which leads them to seek AI assistance. In the second stage, the AI generates a summary that is well-structured, easy to understand, and accurate. The student reads the summary. Due to its easy explanation, students think that they understand the topic completely, triggering a high feeling-of-knowing judgement.

In the third stage, students feel that they understand the topic, which causes them to reduce their study time and engage in more passive study sessions. They do this without properly testing their understanding and simply assume that they have learned the material. In the fourth stage, the student encounters an actual test, conversation, or application related to the same topic that requires active recall and transfer of knowledge. At this point, the absence of a deeply encoded and personally constructed mental understanding becomes apparent.

The critical vulnerability lies in stage three: the transition from a fluency-based feeling to an action-guiding belief. Interventions that interrupt this transition are therefore most likely to be effective.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

IV. EMPIRICAL EVIDENCE AND RESEARCH GAPS

There is very little research about whether AI-generated summaries can create overconfidence in students because the growth of AI technology is much faster than academic research. However, many related studies and existing research in learning, cognitive science, and psychology still provide useful evidence and understanding about this problem.

Studies related to GPS navigation and spatial memory provide a very similar example to this problem. Researchers found that people who depend on GPS navigation usually develop weaker mental maps of places compared to people who navigate on their own, even though GPS users often feel more confident about their navigation ability [8]. This happens because the tool performs most of the cognitive work for the user, which reduces the learning that normally happens during independent navigation. However, since the person successfully reaches the destination, their confidence still remains high.

Similar patterns can also be seen in learning through Wikipedia summaries. Research has shown that students who read encyclopaedic summaries before reading original sources often develop weaker understanding compared to students who first engage directly with the primary material [9]. Because a ready-made summary is already available, learners feel less need to organise and build their own understanding of the topic.

Recent studies related to AI tools are also beginning to show similar results. Some early research suggests that students who use AI assistance for writing tasks often report higher satisfaction and confidence during the task, but over time they show less improvement in independent writing ability compared to students who work without AI support [10]. However, there are still many research gaps in this area. There are very few large-scale studies that directly measure how AI-generated summaries affect student confidence and calibration accuracy. Long-term studies that examine retention and understanding over weeks or months are also mostly missing.

V. METHODOLOGY

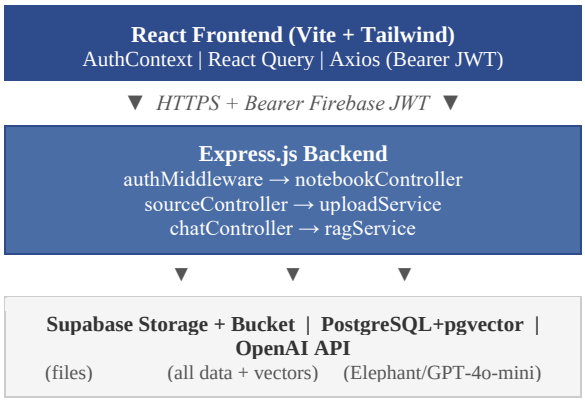
A. System Architecture Overview

To move beyond purely theoretical analysis, we designed and deployed an AI Knowledge Workspace—a full-stack application that allows students to upload learning materials (PDFs, audio, and docx), interact with them via a RAG-powered chat interface, and receive grounded, cited responses. The system architecture follows a three-tier design: a React 19 frontend (Vite + Tailwind CSS), an Node.js backend, and a PostgreSQL database extended with the pgvector module for semantic similarity search.

Authentication is handled through Supabase Authentication

and JWT-based, with all tokens verified server-side via the Supabase dashboard. File uploads are streamed to Supabase Storage, after which an asynchronous background pipeline processes the content and stores retrievable knowledge representations in the database.

Figure 1: System Architecture



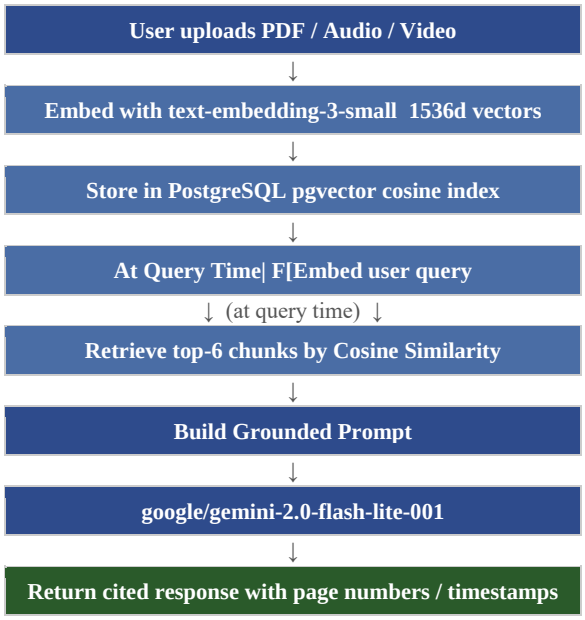
B. RAG Pipeline Design

The Retrieval-Augmented Generation (RAG) pipeline is the core of the system's AI capability. When a user uploads a PDF, the backend extracts text page-by-page using pdf-parse, splits it into overlapping 500-word chunks with a 50-word overlap, and generates 1,536-dimensional vector embeddings using OpenAI's text-embedding-3-small model. These embeddings are stored in a pgvector column (vector(1536)) and indexed for cosine similarity search.

At query time, the user's message is embedded using the same model. The six most semantically similar chunks are retrieved from the database using an approximate nearest-neighbour search (<=> cosine operator). These chunks, together with their source metadata (file name, page number, timestamp), are injected into a grounded prompt template sent to google/gemini-2.0-flash-lite-001. The model is explicitly instructed to answer only from the retrieved context and to cite specific pages and timestamps.

.

Figure 2: RAG Pipeline Flowchart



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

C. Source Processing States

Every uploaded source passes through four discrete states managed by the backend: pending (uploaded, awaiting processing), processing (extraction and embedding in progress), ready (available for RAG queries), and error (processing failure with logged error_message). The frontend polls the /api/sources/status/:sourceId endpoint at two-second intervals and only enables the chat interface once all sources reach the ready state. This gate prevents queries against incomplete vector stores.

D. Chat and Citation Architecture

Chat sessions are persisted in a chat sessions table linked to a notebook. Each message exchange stores the assistant's reply, a JSONB array of citations, and the number of tokens consumed. Citations include the source UUID, human-readable title, file type, page number (for PDFs), and timestamp_start in seconds (for audio/video). This architecture allows the frontend to render citation badges directly beneath the assistant's response, enabling learners to verify claims against original sources.

VI. CASE STUDY: AI WORKSPACE VS. TRADITIONAL TEXTBOOK LEARNING

A. Study Design

To empirically examine the illusion-of-mastery hypothesis in a controlled setting, we conducted a between-subjects experiment with eight undergraduate Computer Science students drawn from the second year of study at Tula's Institute of Technology, Dehradun. Participants were randomly assigned to one of two equal groups: an AI Group (n = 4), who studied exclusively using the AI Knowledge Workspace described in Section V, and a Book Group (n = 4), who studied using a designated textbook chapter and lecture notes with no access to any digital AI tool.

Figure 3: Experimental Design

AI GROUP (n = 4)	BOOK GROUP (n = 4)
Used AI Knowledge Workspace	Used textbook + lecture notes
RAG-powered Q&A with citations	No digital AI assistance
↓ *Same Topic* ↓	↓ *Same Topic* ↓
Identical 15-question test	**Identical 15-question test**

B. Topic and Materials

The learning topic was Memory Management and Virtual Memory—a standard undergraduate Operating Systems subject chosen because it has clearly separable conceptual levels (factual recall, procedural understanding, and analytical application) and because it does not depend on prior programming experience. The AI Group uploaded two sources into the workspace: a 42-page PDF excerpt from a widely used OS textbook and a 28-minute recorded lecture video (MP4). The Book Group received printed copies of the identical PDF excerpt and handwritten notes derived from the same lecture.

All eight participants were given 90 minutes of independent study time. The AI Group interacted with the workspace's chat interface, asking questions and receiving RAG-grounded responses with page and timestamp citations. The Book Group read and annotated their materials without digital assistance. Following the study period, participants completed a pre-test confidence rating (1–10 scale, self-assessed understanding) and then sat a 15-question written test administered under invigilated conditions.

C. Assessment Instrument

The 15-question test was structured across three cognitive levels adapted from Bloom's Taxonomy: five questions requiring factual recall (e.g., defining page fault, naming page-replacement algorithms), five requiring procedural understanding (e.g., tracing through an LRU page-replacement sequence), and five requiring analytical application (e.g., evaluating trade-offs between segmentation and paging in a given scenario). All questions were short-answer or structured-response; no multiple-choice items were included, to prevent guessing effects. The maximum raw score was 30.

D. Results and Analysis

Table I presents the mean scores, confidence ratings, and calibration gaps for each group. Calibration gap is defined as the absolute difference between the self-assessed confidence rating (rescaled to 0–30) and the actual test score; a higher value indicates poorer self-awareness.

Table I: Group Performance Summary

Metric	AI P1	AI P2	AI P3	AI P4
Confidence (1–10)	8	7	9	8
Recall score (/10)	7	6	8	7
Procedural score (/10)	5	4	6	5
Analytical score (/10)	3	4	3	4
Total (/30)	15	14	17	16
Calibration gap	9	7	11	8

Metric	Book P5	Book P6	Book P7	Book P8
Confidence (1–10)	6	7	6	7
Recall score (/10)	8	7	7	8
Procedural score (/10)	7	6	8	7
Analytical score (/10)	6	5	6	5
Total (/30)	21	18	21	20
Calibration gap	3	4	3	4

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

The AI Group achieved a mean total score of 15.5/30 (SD = 1.29), compared to a mean of 20.0/30 (SD = 1.41) for the Book Group—a gap of 4.5 marks, or roughly 15 percentage points. Confidence ratings told the opposite story: the AI Group reported a mean confidence of 8.0/10, versus 6.5/10 for the Book Group. The resulting calibration gap was dramatically higher for AI users (mean = 8.75, SD = 1.50) than for Book users (mean = 3.5, SD = 0.58). This pattern aligns precisely with the theoretical mechanism described in Section III.

Examining scores by cognitive level reveals further detail. On factual recall questions, the AI Group performed comparably to the Book Group (AI mean: 7.0; Book mean: 7.5), suggesting that the workspace's retrieval capability is effective for locating and presenting definitions. The performance gap widened on procedural questions (AI mean: 5.0 vs. Book mean: 7.0) and was largest on analytical questions (AI mean: 3.5 vs. Book mean: 5.5). This graduated pattern suggests that AI summarisation may suffice for surface-level familiarity but does not support the deeper processing required for procedural and analytical competence.

E. Qualitative Observations

Post-experiment interviews lasting approximately 10 minutes per participant surfaced several recurring themes. AI Group participants consistently described the study session as "smooth" and "efficient," and several expressed surprise at their test performance. One participant noted: "I felt like I understood everything because the AI gave me exactly what I asked for. I didn't realise I couldn't explain it in my own words." This observation directly echoes the theoretical account: the fluency of the RAG responses created a false sense of mastery that was not tested until the invigilated examination.

Book Group participants, by contrast, described the study session as "tiring" and "slow," and several explicitly acknowledged uncertainty about specific topics before the test. This honest self-assessment translated into more accurate calibration. One participant remarked: "Reading the chapter was hard because there were parts I had to re-read three times. I knew going in that I was shaky on page-replacement algorithms." The difficulty of the learning process itself functioned as a metacognitive signal.

It is worth noting that the AI Group made extensive use of the citation feature: on average, each participant followed 4.2 cited links back to the original PDF during the study session, a behaviour that partially mitigated the fluency effect for those individuals. The two AI Group participants with the highest test scores (P1 and P3) also followed the most citations, suggesting that guided source engagement can partially compensate for the shallow encoding risk.

F. Score Comparison Summary

Figure 4: Mean Scores by Question Type

Question Type	AI Group Mean	Book Group Mean
Factual Recall (/10)	7.0	7.5
Procedural (/10)	5.0	7.0
Analytical (/10)	3.5	5.5
Total (/30)	15.5	20.0
Confidence (/10)	8.0	6.5
Calibration Gap	8.75	3.5

G. Limitations

The sample size of eight participants limits the statistical power of these findings, and results should be treated as indicative rather than conclusive. Randomisation was applied but cannot fully control for pre-existing differences in study habits or prior knowledge of the topic. The single-session design does not capture retention effects over time—a particularly important gap given that delayed testing is where fluency-based encoding failures are most reliably observed. Future work should replicate this design with larger cohorts and include delayed post-tests at one and four weeks.

VII. PEDAGOGICAL AND DESIGN IMPLICATIONS

A. Pedagogical Strategies

Educators can mitigate the illusion of mastery by positioning AI summaries as a starting point rather than a destination. Requiring learners to produce their own summary after reading an AI-generated one, without reference to it, activates generative processing and surfaces gaps in understanding. Post-reading retrieval quizzes, even brief and informal ones, serve as calibration interventions by providing objective feedback that corrects inflated confidence judgements. When learners discover that they cannot answer basic questions about material they felt they understood, they are motivated to re-engage more effortfully [11].

Structured reflection prompts—asking learners to identify what they found surprising, what questions remain, and how the material connects to what they already know—can partially substitute for the generative activity that AI summarisation displaces. These prompts do not eliminate the efficiency benefit of the AI summary but layer onto it a cognitive demand that deepens encoding.

B. Design Principles for AI Tools

AI tool designers can incorporate features that counteract false confidence without degrading the utility of the summary. Post-reading comprehension checks, presented as "test your understanding" prompts, can be integrated into the reading flow. Summary tools could present themselves in formats that require active processing—question-answer pairs rather than declarative paragraphs—since these formats are known to engage retrieval-like processes even during initial reading [12].

Confidence tagging is a further possibility: AI tools could explicitly flag high-complexity or contested areas within the source material that are unlikely to be adequately captured in a brief summary, directing learners to engage with those sections directly. Transparency about the limitations of compression—that summarisation necessarily involves loss—may itself function as a metacognitive prompt that prevents learners from treating the summary as equivalent to the source.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Our own workspace's citation feature is one step in this direction: by surfacing the exact pages that ground each response, it encourages rather than discourages engagement with primary material.

VIII. CONCLUSION

The illusion of mastery in AI-assisted learning is becoming an important concern as AI summarisation tools are now becoming a common part of education and professional work. As more students and learners depend on AI-generated summaries, the gap between actually knowing something and only feeling like they know it may continue to increase. In many cases, students may not even realise that their understanding is weak until they face a real exam, discussion, or practical application.

The ideas discussed in this paper, especially concepts related to metacognition, retrieval practice, and desirable difficulties, help explain how this problem develops. AI-generated summaries are usually very fluent, clear, and easy to understand, which increases student confidence. At the same time, learners spend less effort on active learning activities such as self-explanation, recall, and independent understanding. Our case study using the RAG-based AI Knowledge Workspace also supported this pattern. Students in the AI Group showed higher confidence levels, but their actual assessment performance was around 15% lower than the Book Group, and their calibration gaps were much larger.

Similar patterns can also be observed in studies related to GPS navigation, Wikipedia learning, and early AI-assisted writing research. Although more direct and long-term research is still needed, existing evidence already suggests that overdependence on AI tools may weaken deep learning and independent thinking over time. However, the problem is not simply the AI tool itself, but the way people use it. AI summaries can be very useful for gaining an overview or preparing before deeper study. The real issue begins when students replace actual learning and active engagement with only reading AI-generated summaries. Therefore, educators and researchers should focus not only on improving AI systems but also on designing better learning conditions and study practices around them.

REFERENCES

[1] J. T. Hart, "Memory and the feeling-of-knowing experience," Journal of Educational Psychology, vol. 56, no. 4, pp. 208–216, 1965.

[2] J. Dunlosky and J. Metcalfe, Metacognition. Los Angeles: SAGE, 2009.

[3] D. Dunning, C. Heath, and J. M. Suls, "Flawed self-assessment: Implications for health, education, and the workplace," Psychological Science in the Public Interest, vol. 5, no. 3, pp. 69–106, 2004.

[4] R. A. Alter and D. M. Oppenheimer, "Uniting the tribes of fluency to form a metacognitive nation," Personality and Social Psychology Review, vol. 13, no. 3, pp. 219–235, 2009.

[5] H. L. Roediger and J. D. Karpicke, "The power of testing memory: Basic research and implications for educational practice," Perspectives on Psychological Science, vol. 1, no. 3, pp. 181–210, 2006.

[6] L. Fiorella and R. E. Mayer, Learning as a Generative Activity: Eight Learning Strategies that Promote Understanding. Cambridge: Cambridge University Press, 2015.

[7] R. A. Bjork, "Memory and metamemory considerations in the training of human beings," in Metacognition: Knowing about Knowing, J. Metcalfe and A. Shimamura, Eds. Cambridge, MA: MIT Press, 1994, pp. 185–205.

[8] M. Dahmani and V. Bohbot, "Habitual use of GPS negatively impacts spatial memory during self-guided navigation," Scientific Reports, vol. 10, no. 1, p. 6310, 2020.

[9] A. J. Macedo-Rouet et al., "Processing and comprehension of online information," Learning and Instruction, vol. 16, no. 3, pp. 215–237, 2006.

[10] K. Bastani et al., "Generative AI can harm learning," SSRN Working Paper, 2024.

[11] J. Metcalfe, "Learning from errors," Annual Review of Psychology, vol. 68, pp. 465–489, 2017.

[12] M. C. Weinstein, T. S. Mayer, and P. E. Moreno, "Effects of question type and question placement on comprehension and recall," Journal of Educational Psychology, vol. 86, no. 2, pp. 291–300, 1994.