Fostering Critical Thinking by Integrating Case-Based Learning and Generative AI in a Closed-Loop Reflective Learning Environment

Hi Kuen Yu[1], Jacky Keung[1], Man On Wong[2], Yicheng Sun[1,*] and Tianqing Xiao[1]
[1]Department of Computer Science, City University of Hong Kong, Hong Kong, China
[2]University of Economics Ho Chi Minh City, Vietnam
hikuenyu2-c@my.cityu.edu.hk, jacky.keung@cityu.edu.hk, wongmanon@ueh.edu.vn, yicsun2-c@my.cityu.edu.hk, shiyaoyan2-c@my.cityu.edu.hk
*Corresponding author

***Abstract*—Critical thinking is a core competency in higher education, particularly in learning contexts involving ill-structured problems, uncertainty, and reasoned decision-making. Although case-based learning can foster higher-order thinking, it often scales poorly and provides limited support for timely, structured assessment of students' reasoning. Large language models (LLMs) offer opportunities for scalable learning support, yet many existing AI-based systems lack pedagogically grounded mechanisms for developing critical thinking. To address these gaps, this study proposes an AI-driven closed-loop learning framework that integrates case-based learning with two specialized LLM agents, a Content Generation Agent (CGA) that produces context-rich scenarios and a Feedback Generation Agent (FGA) that evaluates student reasoning and delivers structured, dimension-specific feedback. A mixed-methods study with 38 postgraduate financial planning students found a statistically significant improvement in domain knowledge and strong performance across critical thinking dimensions, especially evaluation and analysis. Students also reported high engagement and perceived learning effectiveness; however, variability in explanation skills and concerns about feedback timeliness highlight areas for improvement. Overall, the findings suggest that AI-supported closed-loop learning is a promising, scalable approach to developing critical thinking in higher education.**

***Index Terms*—Critical thinking, Case-based learning, Generative AI, Reflective feedback, Personal finance education**

I. Introduction

In higher education, critical thinking develops not through passive knowledge reception, but through repeated cycles of articulating, scrutinizing, and revising one's reasoning [1], [2]. From this perspective, learning is a reflective cognitive process in which learners externalize judgments, receive responses to their thinking, and refine their understanding accordingly [3]. This process, conceptualized as a form of "cognitive echo", is especially important when students confront complex, ill-structured problems with no single predetermined solution [4]. In such contexts, the educational challenge is not simply to help students reach answers, but to support how they reason, justify, and improve those answers over time.

Case-based learning (CBL) offers a strong pedagogical foundation for this kind of reflective reasoning [5]. By engaging students with authentic, context-rich scenarios, CBL prompts them to interpret evidence, analyze competing factors, evaluate alternatives, and justify their decisions [6], [7]. Prior research [8], [9]consistently show that CBL can promote critical thinking by situating learning in realistic problem contexts and requiring active application of knowledge. This is particularly relevant in domains such as personal finance, where decisions often involve uncertainty, trade-offs, and long-term consequences. In such settings, effective learning requires more than factual recall or procedural accuracy; it requires reasoned judgment and reflection on the quality of one's thinking [10].

However, despite its pedagogical strengths, current CBL practice still faces important limitations [5], [8]. First, designing high-quality cases and providing timely, individualized feedback are labor-intensive, making sustained implementation difficult at scale. Second, although CBL aims to develop higher-order thinking, assessment in many case-based settings still emphasizes the correctness of final answers over the quality of the reasoning that produced them. As a result, students may complete case activities without sufficiently structured support to improve specific dimensions of critical thinking [11]. More broadly, many instructional approaches lack a closed-loop mechanism in which student reasoning is elicited, evaluated, responded to, and progressively refined.

Recent advances in large language models (LLMs) offer a promising path for addressing these challenges. LLM-based systems can generate contextualized learning materials, support adaptive interaction, and provide automated feedback at scale [12]. These affordances could reduce the implementation burden of case-based instruction while making reasoning-oriented support more immediate and continuous. However, many current AI-supported educational systems remain oriented toward well-structured tasks, answer generation, or efficiency-driven assistance [13], [14]. Therefore, their capacity to support higher-order reasoning in ill-structured contexts is still limited, and the role of teacher oversight in validating AI-generated content and feedback remains underdeveloped.

These limitations highlight the need for an integrated learning framework that combines the pedagogical strengths of CBL with the scalability of LLM-based support, while still prioritizing reasoning quality and instructional alignment. To meet this need, this study proposes an AI-driven closed-loop learning framework for critical thinking development in higher education and applies it to personal finance, which is a domain that well suited to the study's objectives. Personal finance requires students to develop and apply core knowledge in financial planning, risk, and decision-making, while also grappling with ill-structured problems involving uncertainty, trade-offs, and long-term consequences that demand interpretation, evaluation, justification, and reflection [15]. The framework integrates two specialized LLM-based agents, a Content Generation Agent (CGA) that creates context-rich case scenarios aligned with learning objectives and targeted critical thinking dimensions, and a Feedback Generation Agent (FGA) that evaluates student responses and provides structured, dimension-specific feedback. Teacher oversight is incorporated throughout to ensure pedagogical appropriateness and feedback reliability, enabling an iterative cycle in which student reasoning is continuously elicited, assessed, and refined. Accordingly, this study addresses two research questions:

- **RQ1:** Does the AI-driven learning framework improve students' domain knowledge?
- **RQ2:** How does the framework affect students' critical thinking across key dimensions?

II. Methodology

To meet the need for scalable yet pedagogically grounded support for critical thinking in ill-structured learning contexts, this study proposes an AI-driven closed-loop learning framework. The framework integrates case-based learning, LLM-based support, teacher oversight, and student engagement into a structured instructional cycle.

A. AI-Driven Learning Framework

The framework is implemented through four interconnected phases: Design, Development, Implementation, and Evaluation. In the Design phase, the teacher defines the learning objectives, subject scope, and targeted critical thinking dimensions. During the Development phase, the Content Generation Agent (CGA) produces case-based learning materials, which the teacher then reviews and approves to ensure pedagogical alignment. In the Implementation phase, students engage with the AI-generated scenarios and submit their responses. The Evaluation phase is supported by the Feedback Generation Agent (FGA), which assesses student responses, generates structured feedback, and identifies learning gaps. Based on these evaluation outcomes, the system generates adaptive follow-up tasks targeting the identified weaknesses, forming a continuous feedback loop that iteratively refines both learning content and student reasoning.

Fig. 1 illustrates the framework and the interactions among the teacher, the CGA, the student, and the FGA across these phases. By combining AI-driven content generation with structured feedback and teacher oversight, the framework integrates formative assessment and adaptive learning into a single, coherent process.

B. AI Agent Design

To operationalize the proposed framework, two specialized AI agents: the CGA and the FGA, are designed, with each being responsible for a distinct pedagogical function. The workflow is illustrated in Fig. 2.

1) Content Generation Agent (CGA): The CGA serves as an instructional design component that generates structured case-based learning materials aligned with the intended pedagogical objectives. It operates on teacher-defined inputs, including the subject domain, learning objectives, and targeted critical thinking dimensions. Using these inputs, the CGA explicitly aligns generated scenarios with the specified critical thinking dimensions, ensuring that tasks elicit higher-order reasoning.

To enhance learning effectiveness, the CGA adopts a multimodal content-generation approach with two integrated modules. The textual content generation module produces scenario narratives and reflective questions, while the visual content generation module generates supporting representations that provide contextual cues and facilitate problem interpretation. This multimodal design supports learners' early-stage reasoning by strengthening situational understanding.

During the adaptive learning phase, the CGA generates targeted follow-up cases based on the learning gaps identified. These adaptive materials are designed to strengthen the specific critical thinking dimensions in which students show weaknesses, supporting personalized and iterative learning.

2) Feedback Generation Agent (FGA): The FGA functions as the framework's evaluation and feedback component, converting student responses into structured assessments and actionable guidance. It provides a systematic mechanism for assessing reasoning quality through three stages:

- Dimension mapping to identify critical thinking elements in responses
- Rubric-based scoring using a four-point scale (1–4)
- Generation of structured and evidence-based feedback

The evaluation framework is grounded in five critical thinking dimensions: interpretation, analysis, evaluation, inference and explanation [3], [16]. Feedback is delivered using What Went Well (WWW), Even Better If (EBI), and Situation–Behavior–Impact (SBI) formats, consistent with principles of effective formative and behaviorally grounded feedback [17], [18].

III. Experimental Design

A. Participants

A total of 38 postgraduate students enrolled in a Financial Planning programme participated in the study. All participants had foundational financial knowledge but limited experience with structured critical thinking assessment.

B. Procedure

The study consisted of four stages. First, participants completed a 15-minute pre-test using the Financial Planning Achievement Test (FPAT) to establish baseline knowledge.

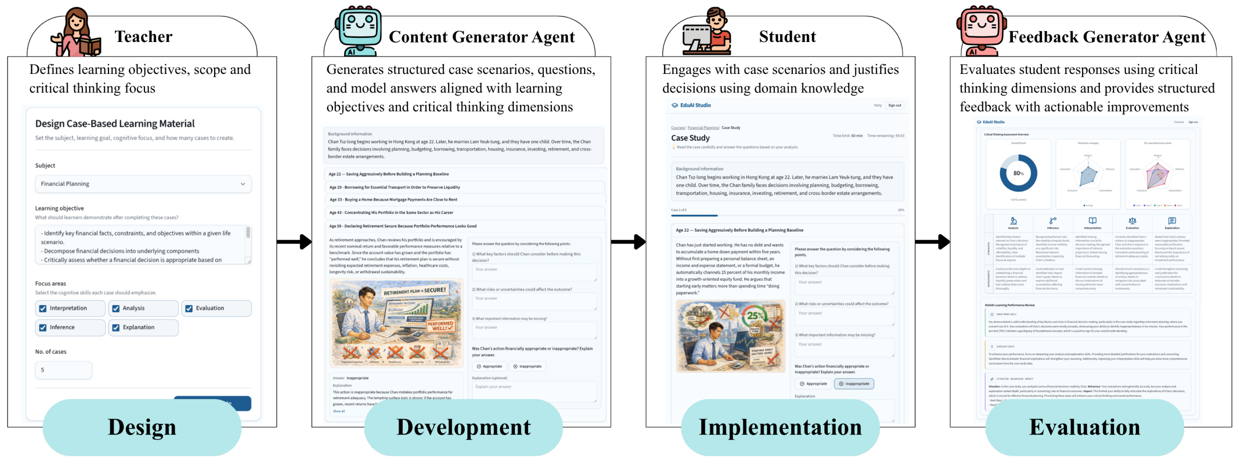

Fig. 1: AI-Driven Closed-Loop Learning Pipeline.

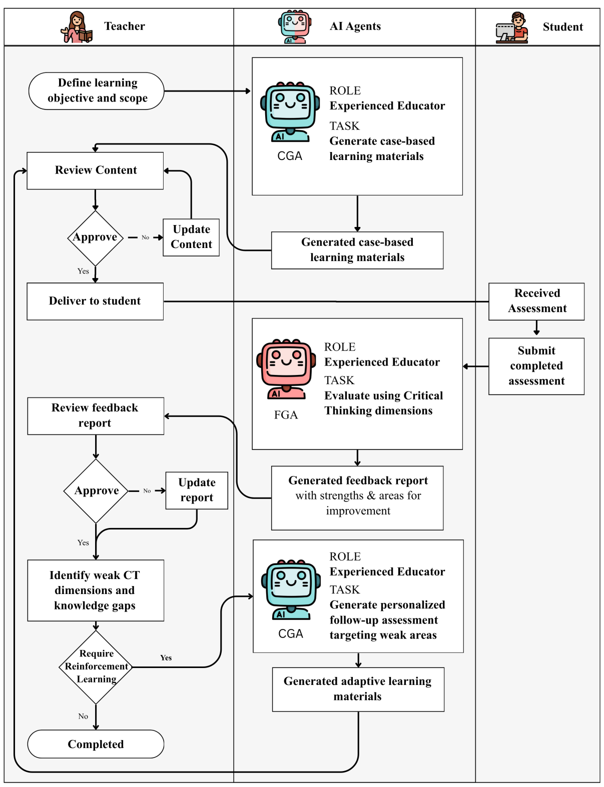

Fig. 2: Workflow on the AI-Driven Closed-Loop Learning Framework.

Second, during a 60-minute intervention, participants worked through AI-generated, case-based scenarios that required analysis, evaluation, and decision justification. Third, participants completed a 15-minute post-test using the same FPAT to measure learning gains. Finally, the FGA evaluated students' case responses, producing dimension-level scores and structured feedback based on a predefined critical thinking rubric. The instructor then verified the evaluation results to ensure accuracy and consistency.

C. Data Collection

Three types of data were collected for quantitative and qualitative analysis: (1) FPAT pre- and post-test scores to measure domain knowledge, (2) written case responses to capture students' reasoning processes, and (3) a post-intervention questionnaire using a five-point Likert scale to assess perceived learning effectiveness, critical thinking development, and engagement. The questionnaire also included an open-ended item to gather feedback on students' learning experiences and suggestions for improvement.

D. Data Analysis

This study employs a convergent parallel mixed-methods design, rooted in a pragmatic research paradigm that prioritizes the research question by using "what works" to address the problem [19], [20].

To evaluate domain knowledge performance, FPAT pre- and post-test scores were analyzed using a paired-sample t-test, and effect size was calculated using Cohen's d to quantify the magnitude of change.

Critical thinking performance was examined using descriptive statistics (means and standard deviations) across five dimensions to identify patterns of competency.

Student perceptions were analyzed using descriptive statistics for Likert-scale items and thematic analysis of open-ended responses. The qualitative analysis followed an inductive approach to identify recurring themes related to learning effectiveness, engagement, and feedback needs.

IV. Results

A. Domain Knowledge Performance

Descriptive statistics showed an increase in FPAT scores from the pre-test (M = 0.77, SD = 0.19) to the post-test (M = 0.82, SD = 0.21), as summarized in Table I. A paired-sample t-test was conducted to assess the effect of the AI-driven intervention on performance. Results indicated a statistically significant improvement, $t(37) = 2.01$, $p = 0.027$, with a mean difference of 0.05. The effect size (Cohen's d = 0.326) suggests a small-to-moderate impact.

TABLE I: Descriptive Statistics for FPAT Scores (N=38)

Test	Mean	Std. Deviation	Min – Max
Pre-test	0.77	0.19	0.30 - 1.00
Post-test	0.82	0.21	0.20 - 1.00

B. Critical Thinking Performance

Critical thinking performance was assessed using students' case-study responses and a five-dimension rubric. Overall performance indicated a relatively high level of competency on a four-point scale (M = 3.37, SD = 0.71). Table II summarizes descriptive statistics for each critical thinking dimension based on students' responses.

TABLE II: Descriptive Statistics of Critical Thinking Dimensions (N=38)

Dimension	Mean	Std. Deviation
Interpretation	3.23	0.66
Analysis	3.32	0.73
Evaluation	3.73	0.41
Inference	3.39	0.66
Explanation	3.18	0.92
Overall	**3.37**	**0.71**

Across the five dimensions, Evaluation had the highest mean score (M = 3.73, SD = 0.41), suggesting that students were generally effective at assessing alternatives and making informed judgments. Inference (M = 3.39, SD = 0.66) and Analysis (M = 3.32, SD = 0.73) also showed strong performance, indicating that students could identify relevant factors and draw reasonable conclusions from the case scenarios.

Interpretation (M = 3.23, SD = 0.66) and Explanation (M = 3.18, SD = 0.92) received comparatively lower scores. Notably, Explanation showed the greatest variability across dimensions, suggesting that students were inconsistent in how clearly they articulated and justified their reasoning. This pattern indicates that, although many students reached appropriate decisions, the clarity and structure of their justifications varied.

These findings suggest that the intervention supported the development of higher-order cognitive skills, particularly evaluative judgment and decision-making. However, additional support may be needed to help students communicate their reasoning more clearly and in a more structured way.

C. Student Perceptions of Learning Experience

Post-intervention questionnaire results showed positive student perceptions across all measured dimensions—learning effectiveness (M = 4.79, SD = 0.52), critical thinking development (M = 4.79, SD = 0.46), and engagement (M = 4.72, SD = 0.55), as summarized in Table III. These consistently high mean scores suggest that students viewed the AI-driven learning approach as both effective and engaging, with particular strength in supporting higher-order cognitive processes.

TABLE III: Student Perceptions of Learning Experience (N=38)

Dimension	Mean	Std. Deviation
Learning Effectiveness	4.79	0.52
Critical Thinking Development	4.79	0.46
Engagement	4.72	0.55

Qualitative responses suggest that students found both the case-based and multiple-choice components beneficial, but for different reasons. Case-based questions were most frequently cited as the most helpful element because they enabled students to apply theoretical knowledge to realistic financial decision-making scenarios, strengthening practical understanding and critical thinking. Several students noted that these scenarios made abstract concepts more concrete and relevant. Multiple-choice questions, in contrast, were valued for reinforcing foundational knowledge by covering key topics, such as liquidity management, retirement planning, and risk control, in a clear and structured manner. A recurring suggestion for improvement was to provide immediate access to correct answers and detailed explanations after submission; without such feedback, students reported limited ability to verify their understanding and learn from mistakes. Overall, the responses suggest that combining theory-based assessment with applied case analysis is effective, but could be strengthened through improved feedback mechanisms.

V. Discussion

This study examined the effectiveness of an AI-driven closed-loop learning framework for supporting domain knowledge acquisition and critical thinking in financial planning education. Overall, the findings suggest that the framework created a structured learning environment in which students could engage with disciplinary knowledge, apply it to realistic cases, and receive targeted support for reasoning development. However, the framework's effectiveness depended not only on the instructional design itself, but also on broader pedagogical conditions, including task structure, feedback timing, and human oversight.

For **RQ1**, the results indicate a statistically significant improvement in domain knowledge, although the effect size was small ($d = 0.326$). This suggests that the intervention supported conceptual learning, but that gains were limited by the short implementation period. This pattern aligns with prior work on AI-supported and case-based learning, which often reports incremental, rather than substantial gains in knowledge improvements from contextualized activities [21]. One possible explanation is that domain knowledge develops through repeated exposure and consolidation over time, whereas the present intervention provided only a brief but intensive experience. Thus, while AI-supported case-based learning may enrich learning tasks, stronger knowledge gains may require more sustained curricular integration.

For **RQ2**, students showed relatively strong performance across the critical thinking dimensions, particularly evaluation, inference, and analysis, suggesting that the case-based design successfully engaged them in the higher-order reasoning required for ill-structured decision-making [5], [6]. The personal finance context may have further supported this outcome by prompting students to weigh alternatives, consider risk, and make judgments under uncertainty. However, the lower and more variable performance in interpretation and explanation point to a more nuanced pattern of development. Although students often arrived at reasonable judgments, they were less consistent in interpreting information early in the task and in articulating the logic behind their decisions clearly. This pattern suggests that critical thinking performance can differ by dimension and may be influenced by task structure and response format. More broadly, it reflects a persistent challenge in higher education, where students may develop evaluative judgment sooner than they develop the ability to communicate their reasoning coherently.

The qualitative findings reinforce this interpretation. Students valued opportunities to apply theory in realistic contexts, suggesting that authenticity and contextual relevance were key drivers of engagement and perceived learning. At the same time, many participants called for immediate access to correct answers and more detailed explanations, underscoring that the effectiveness of AI-supported learning is closely tied to the responsiveness of the feedback environment [22]. Even when case tasks are well designed, delayed or insufficiently transparent feedback can weaken the reflective loop the framework is intended to create. This suggests that successful AI-supported critical thinking interventions depend not only on automated evaluation, but also on how quickly and how clearly learners can use feedback to revise their understanding.

These findings suggest that AI's value in critical thinking education lies less in replacing instruction and more in strengthening the learning environment that supports reasoning. In this study, the framework did more than generate cases or score responses; it created a structured cycle in which students' reasoning was elicited, examined and addressed. This closed-loop structure appears especially important in domains such as personal finance, where learners must navigate ambiguity, competing priorities and long-term consequences. The results also underscore the continuing importance of human involvement: teacher oversight was critical for ensuring the pedagogical appropriateness of AI-generated cases and the reliability of feedback. Thus, in higher-order learning contexts, AI should be framed not as an autonomous instructional substitute, but as part of a human–AI collaborative arrangement.

VI. Conclusion

This study proposed an AI-driven closed-loop learning framework that integrates case-based learning with LLM-based content generation and structured feedback to support critical thinking development in higher education. Applied to personal finance, the framework was designed to meet the need for scalable yet pedagogically grounded support for reasoning in ill-structured problem settings. The findings suggest that the framework can improve domain knowledge, support key dimensions of critical thinking, and foster positive learning experiences. At the same time, the results underscore the importance of timely feedback and human oversight for sustaining an effective reflective learning loop. Given the single-group design and short intervention period, future research should adopt controlled and longitudinal designs to evaluate the framework's effectiveness across broader educational contexts.

References

[1] A. P. Association *et al.*, "A statement of expert consensus for purposes of educational assessment and instruction," *The Delphi Report: Research findings and recommendations prepared for the committee on pre-college philosophy.(ERIC Document Reproduction Service No. ED 315-423)*, 1990.

[2] R. H. Ennis, "Critical thinking across the curriculum: A vision," *Topoi*, vol. 37, no. 1, pp. 165–184, 2018.

[3] P. A. Facione, "Critical thinking: What it is and why it counts," *Insight assessment*, vol. 1, no. 1, pp. 1–23, 2011.

[4] L. Zheng, A. He, C. Qi, H. Zhang, and X. Gu, "Cognitive echo: Enhancing think-aloud protocols with llm-based simulated students," *British Journal of Educational Technology*, vol. 56, no. 5, pp. 2019–2042, 2025.

[5] O. Ten Cate, E. J. Custers, and S. J. Durning, "Principles and practice of case-based clinical reasoning education: a method for preclinical students," 2017.

[6] O. R. Mahdi, I. A. Nassar, and H. A. I. Almuslamani, "The role of using case studies method in improving students' critical thinking skills in higher education." *International Journal of Higher Education*, vol. 9, no. 2, pp. 297–308, 2020.

[7] I. Martins, "Students learn best when they're not handed a map: 4 steps to structure discomfort in your case classroom," 2025.

[8] R. A. Labarrete, "A proposed case-based learning model to operationalize the granular learning theory," *Journal of Education and Training Studies*, 2026. [Online]. Available: https://api.semanticscholar.org/CorpusID:286350089

[9] Y. Xiang, D. Liu, L. Liu, I.-C. Liu, L. Wu, and H. Fan, "Impact of case-based learning on critical thinking dispositions in chinese nursing education: a systematic review and meta-analysis," *Frontiers in medicine*, vol. 12, p. 1452051, 2025.

[10] J. D. Bransford, B. S. Stein, and N. J. Vye, "Helping students learn how to learn from written texts," in *Competent reader, disabled reader*. Routledge, 2024, pp. 141–150.

[11] M. Gallagher and J. Lamb, "Open education in closed-loop systems: Enabling closures and open loops," *Distance Education*, vol. 44, no. 4, pp. 620–636, 2023.

[12] M. Alier, M. J. Casañ, and D. A. Filvà, "Smart learning applications: leveraging llms for contextualized and ethical educational technology," in *International conference on technological ecosystems for enhancing multiculturality*. Springer, 2023, pp. 190–199.

[13] A. M. Basiouni, M. El Rashid, and K. Shaalan, "In-context learning in large language models (llms): Mechanisms, capabilities, and implications for advanced knowledge representation and reasoning," *IEEE Access*, 2025.

[14] G. Yadav, Y.-J. Tseng, and X. Ni, "Contextualizing problems to student interests at scale in intelligent tutoring system using large language models," *arXiv preprint arXiv:2306.00190*, 2023.

[15] A. Lusardi and O. S. Mitchell, "The economic importance of financial literacy: Theory and evidence," *Journal of economic literature*, vol. 52, no. 1, pp. 5–44, 2014.

[16] P. A. Facione, "The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill," *Informal logic*, vol. 20, no. 1, 2000.

[17] J. Hattie and H. Timperley, "The power of feedback," *Review of educational research*, vol. 77, no. 1, pp. 81–112, 2007.

[18] A. N. Kluger and A. DeNisi, "The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory." *Psychological bulletin*, vol. 119, no. 2, p. 254, 1996.

[19] V. Kaushik and C. A. Walsh, "Pragmatism as a research paradigm and its implications for social work research," *Social sciences*, vol. 8, no. 9, p. 255, 2019.

[20] J. L. Manion, *A mixed methods investigation of student achievement and satisfaction in traditional versus online learning environments*. Lindenwood University, 2019.

[21] D. N. C. Dey, "Enhancing educational tools through artificial intelligence in perspective of need of ai," *Available at SSRN 5031275*, 2024.

[22] R. L. Amyatun and A. Kholis, "Can artificial intelligence (ai) like quillbot ai assist students' writing skills? assisting learning to write texts using ai," *ELE Reviews: English Language Education Reviews*, vol. 3, no. 2, pp. 135–154, 2023.