Kashihara, A. et al. (Eds.) (2024). Proceedings of the 32nd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Supporting Students' Post-Exam Reflection Needs in College Automation Engineering Course Using LLM

Edward ANOLIEFO[a*], Patrick OCHEJA[b], Regina OCHONU[c], Brendan FLANAGAN[b] & Hiroaki OGATA[b]

[a]*Department of Electronic Engineering, University of Nigeria, Nsukka, Nigeria*
[b]*Academic Centre for Computing and Media Studies, Kyoto University, Japan*
[c]*Dept. of Signal Theory and Communications, Universitat Politècnica de Catalunya, Spain*
*ocheja.ileanwa.65s@st.kyoto-u.ac.jp

Abstract: Post-exam reflection is critical in helping students consolidate knowledge acquired during a course, enabling them to apply this understanding in future professional contexts. This study investigates the effectiveness of Mirai, a large language model-based (LLM) chatbot, in supporting students' post-exam reflection needs in an Automation Engineering course. Through a controlled experiment, we explored how context-tuned and non-context-tuned versions of Mirai impacted students' reflection habits, help-seeking behaviors, and perceptions of the tool. Students interact with the chatbot to clarify exam questions and receive personalized explanations. A a surveys based on the extended technology acceptance model (exTAM) was conducted and the resulting data was analyzed. We assessed the efficacy of Mirai in facilitating a deeper understanding of exam-related material, improving students' knowledge, engagement and performance. The findings from this study provide insights into the immediate educational benefits of LLM-based tools, their acceptance among students, and their role in enhancing learning outcomes in engineering education.

Keywords: Post-exam Reflection, Large Language Models (LLMs), Automation Engineering Education, Context-Tuned AI, Student Engagement, Technology Acceptance Model (TAM), AI-Driven Learning Tools

1. Introduction

Large language models (LLMs), represent a significant advancement in natural language processing (NLP), enabling machines to understand and generate human-like text. These models, built on deep learning architectures, are trained on vast amounts of data, allowing them to generate contextually relevant and coherent responses across a wide range of topics (Brown, Mann, Ryder, et al., 2020). The versatility and adaptive capabilities of LLMs have made them indispensable in various applications, including education, where they have the potential to revolutionize traditional teaching and learning practices. In educational settings, particularly in fields that demand a high level of technical precision like Automation Engineering, the ability to interact with and leverage such advanced AI systems can be transformative. LLMs can facilitate personalized learning experiences by providing immediate feedback, answering complex questions, and engaging students in meaningful dialogue about their coursework. This ability to simulate human-like interactions with AI presents new opportunities for supporting reflective practices, which are critical in helping students consolidate their learning and apply theoretical knowledge in practical contexts (Roll & Wylie, 2016).

Reflection, especially after examinations, is a key metacognitive process that enables students to critically assess their understanding, identify knowledge gaps, and reinforce their learning (Moon, 2013). In Automation Engineering, where students must grasp intricate

concepts and apply them in both academic and professional settings, post-exam reflection becomes particularly important (Socha, Razmov & Davis, 2003). However, traditional reflection methods often fall short, as they can lack immediacy, personalized feedback, and the ability to engage students actively (Yoder & Chenoweth, 2020). LLMs, such as Gemini (Team, Anil, Borgeaud, et al., 2023) can address these challenges by providing an interactive and adaptable platform for post-exam reflection. Gemini can help students revisit and clarify difficult concepts encountered during exams, thereby supporting deeper engagement with the material and enhancing overall learning outcomes. Thus, we set out to answer the following research questions:

1. How effective is Mirai, a conversational chatbot, in supporting students' post-exam reflection needs in Automation Engineering?
2. What are immediate gains, effects and students' perception about LLM-based chatbot in Automation Engineering?

This study explores the effectiveness of Mirai in facilitating post-exam reflection among Automation Engineering students, examining its impact on help-seeking behaviors, learning outcomes, and student perceptions of AI-based learning tools.

2. Related work

Reflection has long been recognized as a critical component of effective learning, particularly in higher education. Schön's seminal work on reflective practice (Schön, 2017), laid the foundation for understanding how reflection helps bridge the gap between theoretical knowledge and practical application. This concept has been expanded by Moon (2013), who argued that structured reflection can lead to deeper learning and better retention of complex concepts. These theories have been instrumental in shaping educational practices that encourage students to engage in reflective activities post-exam. In the context of engineering education, reflection is particularly important due to the complex and often abstract nature of the material. Litzinger, Wise and Lee (2005) demonstrated that reflective practices could significantly enhance students' problem-solving abilities and their understanding of core engineering concepts. Moreover, Ash and Clayton (2009) have shown that structured reflection can improve students' critical thinking skills, which are essential for success in engineering disciplines.

The rise of AI and LLMs has opened new avenues for supporting reflective practices. Zawacki-Richter, Marín, Bond, & Gouverneur (2019) presented a comprehensive review of AI applications in education, highlighting the potential of AI to offer personalized learning experiences. Similarly, Roll and Wylie (2016) discussed the role of AI in facilitating metacognitive activities, such as reflection, by providing immediate and context-sensitive feedback. However, there is a paucity of research specifically addressing the use of LLMs in post-exam reflection, particularly in technical fields like Automation Engineering. While some studies have investigated the use of AI in general education settings, there is a need for more targeted research that examines its impact on specific courses and disciplines (Abedi, Alshybani, Shahadat, & Murillo, 2023). This study seeks to fill this gap by examining the effectiveness of Mirai, an LLM-based tool, in supporting students' post-exam reflection in an Automation Engineering course for final year college electronic engineering students. By building on the existing literature and exploring the application of LLMs in a new context, this research contributes to the ongoing discussion on how to enhance student learning outcomes through innovative pedagogical approaches.

3. Methodology

3.1 System Design

In this study, we use the system called Mirai, a learning analytics and LLM-powered chatbot illustrated in Figure 1. Mirai integrates multiple components to facilitate interactions with

students. The primary interface allows students to initiate a conversation with Mirai by selecting from predefined options such as "I want more explanation about the question," "I have a specific question on this question," or "I am not sure what to ask". These options guide the AI's response, directing it to either provide a deeper explanation, address specific queries, or offer suggestions for further inquiry.

However, not all elements of the system depicted were utilized in the current study. Notably, the components labeled "Explainer" and "Suggestor"—which provide detailed explanations and suggest relevant follow-up questions—were excluded from this study. Instead, the focus was solely on the core interaction between the student and the context-tuned AI, which was enhanced with specific lecture notes relevant to the course. We hypothesize that this context-tuning could improve the AI's ability to provide personalized, contextually relevant feedback, thereby enhancing the reflective learning process. The omission of the Explainer and Suggestor allowed for a more controlled investigation into the effects of context-tuning on students' post-exam reflection and learning outcomes.

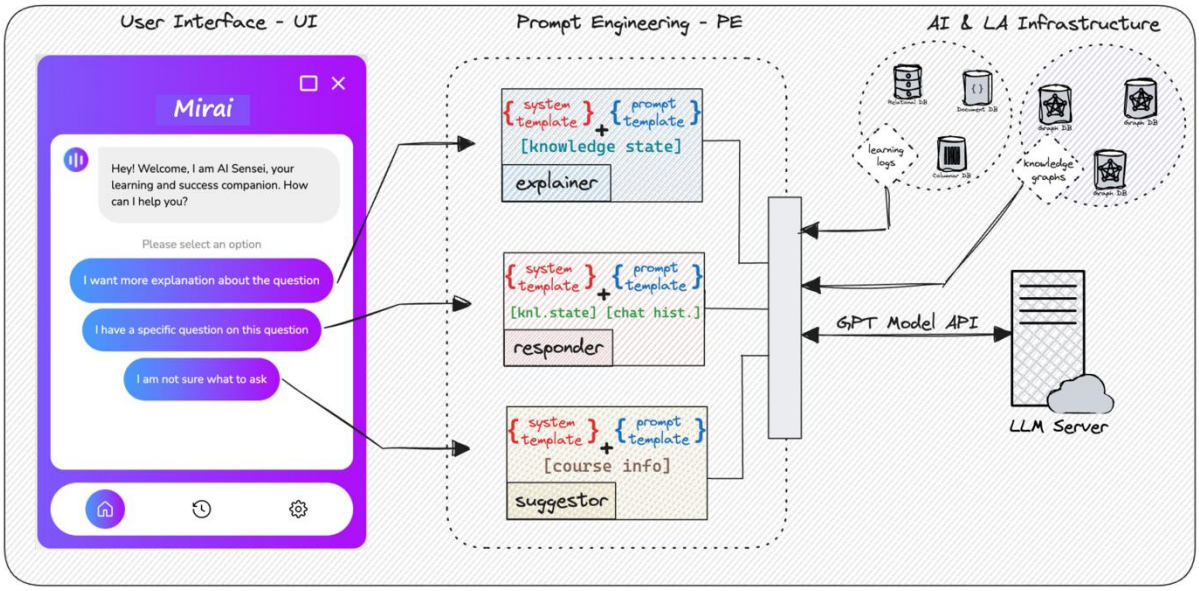

Figure 1. Mirai AI chatbot system overview.

3.2 Study Design

This study employs a mixed-methods approach to evaluate the effectiveness of Mirai in supporting post-exam reflection among students in a college-level Automation Engineering course. Participants were randomly assigned to either a control group or an experimental group. The control group interacted with a standard version of Mirai (no context-tuning), while the experimental group used a context-tuned version, customized with specific lecture notes from the course. The goal is to evaluate the impact of context-tuning on students' post-exam reflection habits, help-seeking behaviors, and overall learning outcomes. Participants were first asked to select a specific question from their recent exam that they found challenging or confusing. They then engaged in an interactive chat with Mirai, asking questions and seeking clarifications related to the selected exam question. The interaction was designed to mimic a reflective dialogue, encouraging students to critically engage with the material and seek deeper understanding.

After the chat, students completed a questionnaire based on the extended Technology Acceptance Model (exTAM) instrument (Wu & Gao, 2011), which assesses various dimensions of their interaction with the AI, including perceived usefulness, ease of use, and overall satisfaction (Granić & Marangunić, 2019). The questionnaire administered comprises 17 mixed-format questions where respondents express their level of agreement with statements on a seven-point Likert scale, ranging from Strongly Disagree to Strongly Agree.

The scale items are averaged across five key factors: Perceived Ease of Use (PEOU), Perceived Usefulness (PU), Perceived Enjoyment (PE), Attitude (ATT), and Intention.

4. Results

Table 1 presents the descriptive statistics of the extended Technology Acceptance Model (exTAM) factors for both the control and experimental groups. The table compares the mean scores and standard deviations across five factors: Perceived Ease of Use (PEOU), Perceived Usefulness (PU), Perceived Enjoyment (PE), Attitude (ATT), and Intention. The experimental group, which interacted with a context-tuned version of the AI tool, showed slightly lower mean scores in PEOU compared to the control group, indicating that students found the context-tuned tool slightly less easy to use. However, the experimental group reported higher mean scores in Perceived Usefulness, Perceived Enjoyment, and Intention, suggesting that despite the slight reduction in ease of use, the context-tuned tool was perceived as more useful and enjoyable, with a higher intention to use the tool in the future. The reliability of the factors, indicated by Cronbach's alpha, is generally high, supporting the consistency of the measurements, except for Attitude, which shows moderate reliability.

Table 1. *exTAM Descriptive Results*

Groups	**Contr.**	**Exp.**	**Contr.**	**Exp.**	**Contr.**	**Exp.**
Number of Samples	21	28	21	28	21	28
Factor	**Reliability, α**		**Mean, μ**		**Std. Dev., σ**	
Perceived Ease of Use (PEOU):	0.95	0.97	6.33	5.98	1.07	1.34
Perceived Usefulness (PU):	0.96	0.96	5.68	5.74	1.26	1.57
Perceived Enjoyment (PE):	0.96	0.96	4.73	5.52	1.70	1.70
Attitude (ATT):	0.57	0.62	5.56	5.56	1.32	1.47
Intention:	0.72	0.90	5.06	5.67	1.30	1.66

The Mann-Whitney U-tests were conducted to compare the control and experimental groups across the five exTAM factors, as the assumptions of normality and homogeneity of variance required for a t-test were not met. The results indicate that there were no significant differences between the two groups in PEOU, PU, and ATT, with p-values greater than 0.05. However, a significant difference was observed in Perceived Enjoyment (PE) and Intention, with p-values of 0.041 and 0.019, respectively. The effect sizes for these factors were small to moderate, indicating that while the differences are statistically significant, they may not be practically substantial.

5. Discussion

This study aimed to explore the effectiveness of the Mirai in supporting students' post-exam reflection needs in an Automation Engineering course, focusing on how context-tuned and non-context-tuned versions of the AI tool impacted students' learning gains, reflection habits, help-seeking behaviors, and overall perceptions. The results from the exTAM-based questionnaire, as well as the Mann-Whitney U-test analyses, provide nuanced insights into the efficacy of these AI-driven interventions in enhancing learning outcomes. Reflecting on the research questions posed, several key observations can be made.

Firstly, regarding the effectiveness of Mirai in supporting post-exam reflection, the results demonstrate mixed outcomes. Although the experimental group, which interacted with the context-tuned version of the LLM, reported higher perceived usefulness (PU) and enjoyment (PE), the differences in ease of use (PEOU) were slightly lower compared to the control group. This suggests that while the additional contextualization may have enhanced

the tool's relevance and engagement, it may have introduced complexity that made it slightly less user-friendly. The significant increase in perceived enjoyment and intention to use the tool indicates that students valued the tailored feedback provided by the context-tuned Mirai, which likely contributed to a more meaningful and satisfying reflection experience. These findings align with previous literature that emphasizes the importance of personalization in educational technologies for fostering deeper engagement and learning.

Secondly, the analysis of help-seeking behaviors using the Mann-Whitney U-test reveals that while there were no significant differences in most factors (PEOU, PU, ATT) between the two groups, the experimental group showed significantly higher scores in PE and intention to use the tool in the future. This suggests that the context-tuned version of Mirai may have better supported students' reflective practices by making the interaction more enjoyable and increasing their likelihood of engaging with the tool in the future. The lack of significant differences in PEOU and ATT, however, highlights the need for balancing contextual depth with ease of use to ensure that educational tools remain accessible while still offering personalized support. The small to moderate effect sizes observed in PE and intention also suggest that while the improvements are statistically significant, they may not translate to dramatic changes in behavior or perception, indicating an area for further refinement in future iterations of the tool.

Analyzing the feedback from students provided deeper insights into their perceptions of the LLM-based chatbot. In the control group, comments were sparse and mostly neutral, with some expressing appreciation for the tool's interactivity but also noting areas where the system did not fully meet expectations, such as not always providing optimal results. In contrast, the experimental group provided more extensive and diverse feedback. Students highlighted the chatbot's effectiveness in making the reflection process enjoyable and interactive, with many appreciating its human-like responses and the guidance it provided in problem-solving. Some comments emphasized the tool's potential to be scaled for broader use across different courses, reflecting a strong perceived value in the AI's ability to support learning beyond the scope of the study. In the experimental group, students praised the chatbot's speed, and ability to provide well-explained, detailed answers that enhanced their understanding of complex concepts. These students found the chatbot to be user-friendly and appreciated its ability to challenge them with follow-up questions, which they felt improved their critical thinking and problem-solving skills. This suggests that the immediate effects of using the LLM-based chatbot were largely beneficial, contributing to both cognitive and affective gains in learning. The control group, while also recognizing the chatbot's utility, provided less enthusiastic feedback, indicating that the additional context provided in the experimental version significantly enhanced the user experience. These findings demonstrate that context-tuning is a crucial feature for maximizing the educational impact of AI-driven tools in engineering education settings.

6. Conclusion

In conclusion, this study demonstrated that the Mirai can support students' post-exam reflection needs in Automation Engineering by providing personalized, context-tuned feedback, which was particularly well-received by the experimental group. The analysis revealed that while the tool enhanced perceived usefulness and enjoyment, slight reductions in ease of use were observed, suggesting a need for further refinement. Immediate gains were evident in the students' engagement, with many appreciating the AI's human-like interaction and problem-solving guidance, though the control group indicated a less enthusiastic response. The study's limitations include the relatively small sample size and the focus on a single technical discipline, suggesting the need for future research to explore the scalability of such AI tools across diverse educational settings and larger cohorts. Future work should also address the balance between usability and contextual depth to further enhance the educational impact of LLM-based tools.

Acknowledgements

This work was supported in part by the following grants: Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Fellows JP22KJ1914, (B) JP20H01722, JP23H01001, (Exploratory) JP21K19824, (B) JP22H03902, New Energy and Industrial Technology Development Organization (NEDO) under Grant JPNP20006 and Grant JPNP18013 and Google Cloud Platform (GCP) Credit Award - PaliGemma Academic Program.

References

Abedi, M., Alshybani, I., Shahadat, M. R. B., & Murillo, M. (2023). Beyond Traditional Teaching: The Potential of Large Language Models and Chatbots in Graduate Engineering Education. *Qeios*.

Ash, S. L., & Clayton, P. H. (2009). Learning through critical reflection: A tutorial for students in service-learning (Instructor version). *Raleigh, NC: Authors*.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada*

Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context: A systematic literature review. *British Journal of Educational Technology*, *50*(5), 2572-2593.

Litzinger, T. A., Wise, J. C., & Lee, S. H. (2005). Self-directed learning readiness among engineering undergraduate students. *Journal of Engineering Education*, *94*(2), 215-221.

Moon, J. A. (2013). *Reflection in learning and professional development: Theory and practice*. Routledge.

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, *26*, 582-599.

Schön, D. A. (2017). *The reflective practitioner: How professionals think in action*. Routledge.

Socha, D., Razmov, V., & Davis, E. (2003). Teaching reflective skills in an engineering course. In *2003 Annual Conference* (pp. 8-1080).

Team, G., Anil, R., Borgeaud, S., Wu, Y., Alayrac, J. B., Yu, J., ... & Ahn, J. (2023). Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Wu, X., & Gao, Y. (2011). Applying the extended technology acceptance model to the use of clickers in student learning: Some evidence from macroeconomics classes. *American Journal of Business Education*, 4, 43–50.

Yoder, J. A., & Chenoweth, S. (2020). Reflection for recovery: exam wrappers in an object-oriented software development course help struggling students improve future exam scores. In *2020 IEEE 32nd Conference on Software Engineering Education and Training (CSEE&T)* (pp. 1-9). IEEE.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International Journal of Educational Technology in Higher Education*, *16*(1), 1-27.