

The Possibility of Using Generative AI Tools for English Speaking

JINGUSHI, Kenta
(Kansai University)

1. Introduction

In recent years, the advancement of Large Language Models (LLMs) has greatly broadened the possibilities for incorporating AI into language learning environments (Brown et al., 2020). AI chatbots have been recognized for its strong potential to deliver individualized learning experiences for English language learners (Zhai et al., 2023). Remarkably, ChatGPT attracted over one million users within its first four days, and by January 2023, its user base had surpassed 100 million.

According to Student Generative AI Survey 2025 (Josh Freeman, 2025) comparing AI usage among university students in 2024 and 2025, the most common application of AI was generating text (e.g., ChatGPT), with usage rising significantly from 30% in 2024 to 64% in 2025. Other widely adopted uses included enhancing and editing writing (e.g., Grammarly). In contrast, language-related applications such as translation tools (e.g., Google Translate) and speech-to-text transcription (e.g., YouTube captions) were used less frequently—35% and 24% respectively in 2025. These findings suggest that university students primarily use AI for academic productivity rather than for language learning especially improving speaking skill.

This paper explores the potential of generative AI tools in supporting English as a Foreign Language (EFL) learners, with a specific focus on how these tools can be utilized to enhance speaking skills. While recent advancements in LLMs have opened new avenues for AI-assisted language instruction, the effectiveness of such tools largely depends on how learners and educators interact with them. In this regard, prompt engineering has emerged as a vital strategy for optimizing the performance of LLMs (Liu et al., 2021), allowing for more tailored and goal-oriented language practice.

Despite the growing interest in generative AI within the field of education, research on its practical application in EFL speaking contexts—especially from the perspective of prompt design—remains limited. Building on the theoretical insights of Reynolds and McDonough (2021), this study presents a methodological framework for constructing prompts that align with learners' proficiency levels and speaking objectives. By examining how different prompt strategies influence learner interaction and

motivation, this research seeks to contribute to the development of more effective, learner-centered approaches to EFL speaking instruction using generative AI technologies. Ultimately, the study aims to broaden our understanding of how AI can facilitate personalized, meaningful language practice in educational settings.

2. Literature Review

The incorporation of generative AI tools like ChatGPT into EFL education has been attracting growing interest (Han, Jieun, et al., 2023); nonetheless, studies focusing on the role of prompt engineering in EFL settings are still scarce. Zheng, Liu, et al. (2024) explored the adoption of generative AI in EFL learning environments, with particular attention to how learners' motivation affects their acceptance of such technologies. Their research utilized the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) in combination with Self-Determination Theory (Deci & Ryan, 2013). . Furthermore, the study emphasized that to effectively foster learner motivation when applying generative AI in EFL instruction, it is essential not only to implement the technology itself but also to establish a learning environment that nurtures learners' autonomy and facilitates the attainment of their educational goals.

In the field of reading comprehension, the ability to automatically generate high-quality questions has long been a desirable goal, given the significant time and effort educators must dedicate to crafting them. Nonetheless, persistent issues such as inconsistent question quality and the lack of transparency and validation in evaluation procedures have posed considerable challenges (Kurdi, 2020). In response, Lee, Unggi, et al. (2024) conducted a study utilizing Design and Development Research to build and validate an Automatic Question Generation (AQG) system powered by ChatGPT's LLMs.

In a related study, Young & Shishido (2023) investigated how ChatGPT could be used to generate sample dialogues for EFL chatbot systems, aiming to address the problem that many EFL learners have limited chances to engage in real-life conversations, which hampers their speaking development. Their findings revealed that the dialogues created were particularly appropriate for A2-level learners, as the content was accessible and utilized familiar vocabulary. Furthermore, they noted that most of the B1-level dialogues effectively introduced new vocabulary and stimulated learning. The researchers emphasized the importance of experimenting with various prompt engineering strategies to improve both the diversity and the pedagogical quality of the AI-generated conversations.

Ali et al. (2023) introduced the concept of "in-context learning" in developing TeacherGAIA, a generative AI chatbot. By embedding detailed instructions and examples in prompts, their chatbot was able to better understand context and perform complex tasks. This approach demonstrated how AI-powered tools can facilitate

learner autonomy and self-assessment when supported by effective prompt engineering strategies.

3. Problem Statement

As highlighted in the preceding literature review, a growing body of research underscores the transformative potential of generative AI technologies in EFL education. However, maximizing the effectiveness of these tools requires thoughtful integration into pedagogical practice, particularly through the creation of appropriate prompts tailored to learners' proficiency levels. While generative AI tools are capable of generating a wide range of outputs suitable for various English proficiency levels, research has shown that the effectiveness of these outputs greatly depends on how well the prompts align with learners' needs. For instance, simpler and more concrete prompts tend to be more beneficial for beginner learners, whereas abstract and complex prompts are better suited for advanced learners (Young & Shishido, 2023). Designing such level-appropriate prompts necessitates both pedagogical knowledge and practical teaching experience.

The emergence of generative AI technologies is undoubtedly transforming the role of EFL educators. Instead of serving solely as providers of knowledge, teachers are now increasingly positioned as facilitators who support students through personalized guidance and adaptive teaching strategies. As noted by Nitta & Baba (2015), EFL instruction has been evolving from a learner-centered to a more person-centered approach. To fully realize the benefits of generative AI as a pedagogical tool, a new educational framework is needed—one that encourages collaboration between teachers and AI systems. In this context, teacher-guided prompt engineering may become a key mechanism for integrating generative AI into language learning environments. Such collaboration holds promise for enhancing learner engagement, promoting individualized instruction, and alleviating the workload of educators. As the educational landscape continues to evolve, developing prompt engineering competencies is likely to become an essential skill set for EFL teachers.

4. Purpose of the Study

The purpose of this study is to explore how generative AI tools can be effectively utilized to facilitate English speaking practice for EFL learners through the use of prompts informed by prompt engineering theory (Zero shot and Chain-of-thought prompting). In particular, this research examines how prompts that are carefully designed to reflect learners' proficiency levels and learning objectives influence the quality and educational relevance of AI-generated conversational exchanges. As previous studies have noted, the quality of content produced by generative AI models is significantly shaped by the design of the prompts they receive. Therefore, prompt engineering plays a vital role in optimizing the interaction between learners and AI,

especially in the context of EFL speaking education.

To provide meaningful and effective speaking practice, generative AI tools must generate conversations that are appropriately tailored to the learners' language levels and communicative goals.

In this context, the teacher's role is to design prompts that are aligned with pedagogical aims and to guide the structure and flow of learner-AI interactions. If the conversations generated by AI do not support the intended learning outcomes or lesson framework, their educational value may be limited. Thus, it is essential for educators to apply their professional expertise in creating prompts that reflect learners' specific needs and goals. By doing so, generative AI can be leveraged not just for conversational practice, but as a means to deepen learners' engagement and learning. This study aims to investigate how teacher-facilitated prompt engineering can enhance the use of generative AI in supporting English speaking development for EFL learners.

RQ1:

Is there a difference in the number of words spoken by learners when using two different types of prompts?

This research question aims to investigate how the design of prompts—such as their complexity, specificity, or instructional clarity—affects the amount of spoken output from EFL learners during interactions with generative AI tools. By comparing the number of words produced under different prompting conditions, the study seeks to determine whether prompt design plays a significant role in encouraging more extensive language use. The results may provide insights into how to structure prompts to maximize speaking practice through AI interaction.

RQ2:

How can generative AI tools influence learners' motivation in English speaking practice?

This question explores the motivational impact of using generative AI tools in English speaking tasks. It focuses on understanding how learners perceive and respond emotionally and cognitively to engaging in conversations with AI—whether the experience increases their willingness to speak, reduces anxiety, or fosters a more autonomous and enjoyable learning environment. By analyzing learner feedback and behavioral responses, the study aims to assess how AI can serve as a motivational support in language education.

5. Prompt Engineering

A prompt is an instruction or input that guides the behavior of LLMs, such as those powered by generative AI. The effectiveness of the model's response is significantly influenced by how well the prompt is formulated (Shishido, 2024). The method of constructing such prompts is referred to as prompt engineering, and its educational applications—particularly in AI-assisted language learning—are receiving increasing scholarly attention (Young & Shishido, 2023). This study outlines three fundamental prompt engineering strategies to explore how generative AI tools can be effectively utilized in English speaking contexts.

Zero-shot prompting involves asking the AI to perform a task without providing any specific examples. In contrast to few-shot prompting, which includes examples, zero-shot prompting relies solely on the model's internal knowledge and generalization ability. Li (2023) identifies key benefits of this technique, including its efficiency—because no prior examples are required, making it faster and more cost-effective; its flexibility—due to its adaptability across a wide range of tasks; and its ability to reduce bias, as it avoids the influence of specific data patterns.

Few-shot prompting provides the AI with a limited number of task-specific examples before requesting a response. This technique takes advantage of the model's capacity to infer task patterns from the examples and apply them effectively. It has been widely applied to tasks like language translation, question answering, and text summarization (Reynolds, L., & McDonell, K., 2021), and may also contribute to more natural spoken language output when designed appropriately.

Chain-of-thought prompting encourages the model to reason through problems in a step-by-step manner, resembling human cognitive processes. By including explicit reasoning cues like “Let's think step by step,” the model can produce more structured and thoughtful responses. Wei et al. (2022) demonstrated that this approach enhances the model's reasoning capabilities, improving the depth and clarity of generated output.

6. Materials

In this study, I reused the data originally collected for the study *Prompting Relationship: Using AI for EFL Motivation* (Jingushi, 2024b), which focused on the interaction between learners and AI tools. Specifically, the data consist of recorded conversations generated by a generative AI tool in response to different types of prompts based on prompt engineering techniques.

The AI tool used was ChatGPT for iOS version 1.2024.115 (free plan), with the

language setting on auto-detect, no personalization settings applied, and memory function enabled. Data were collected between October 1 and November 6, 2024.

To ensure consistency and accuracy in prompt usage, each prompt was sent to the participants by the researcher via text, and participants were instructed to copy and paste the exact prompts into the AI chat interface. Since ChatGPT retains prior conversation history, each prompt type was initiated as a separate new conversation to avoid contextual influence.

Four participants (A, B, C and D), all business professionals learning English, took part in the study. Participant A regularly uses English in both professional and personal contexts, reports no anxiety using the language, and has an advanced proficiency level (CEFR C1). Participant B rarely uses English in daily life but frequently takes standardized English tests such as TOEIC, Eiken, and IELTS to maintain motivation; his proficiency level is approximately CEFR B2. Participant C works for a foreign-based company where English is the common language (approximately B2-C1 level). He uses ChatGPT daily to check his English expressions and has incorporated AI tools into his language learning practice. Lastly, Participant D is an English teacher in a high school. She is also enthusiastic and struggle to develop her English; her English proficiency level is around CEFR B2 as well.

I hypothesize that learners at higher proficiency levels (e.g., B2–C1 on the CEFR scale) can serve as effective exploratory case studies to investigate the interactional and motivational potential of AI-assisted language learning. Moreover, their experience can help generate insights into optimal prompting techniques, perceived benefits, and motivational shifts—insights that can be strategically adapted, simplified, and scaffolded for lower-proficiency learners (A2–B1).

The contents of the materials

(1) Free conversation in English with AI for 5 minutes ✖ 3 days

In this condition, participants initiated English conversations with ChatGPT without any guidance or constraints. One participant began the dialogue simply with "Hello," while another used the phrase "Can you speak English, please?" to start the interaction.

(2) English conversation with AI at prompts specified by the researcher: ✖ 3 days

This prompt was deliberately designed to elicit both conversational output and feedback from the AI, allowing participants to gain language input and reflection opportunities simultaneously. By comparing these two conditions—free-form and prompt-guided—the study aimed to explore how tailored prompting can enhance the

quality and motivational aspects of AI-assisted learning, highlighting the practical implications of prompt engineering for EFL learners.

(3) Web questionnaire 5 minutes

To evaluate participants' experiences and perceptions, I conducted a short web-based questionnaire after the above (1) and (2). The survey included both multiple-choice and open-ended items as shown in the Appendices.

(4) Zoom interview to supplement the above survey 10 minutes

Following the questionnaire, a brief semi-structured interview was conducted via Zoom to gather more nuanced insights into participants' experiences.

7. Results

As Jingushi (2024b) mentioned, the word counts were recorded for the prompt for the content of (1) and (2) as P1 and P2, and Day1, Day2, and Day3, respectively. In reference to the labels used in Figure 1 (Jingushi, 2024b), "ChatGPT-A" denotes the total number of words generated by the AI during interactions with Participant A. The notation "P1-Day1" indicates the word count produced on the first day using the prompt identified as (1). Additionally, for the survey results described in section (3), each participant (A, B, and C) provided responses on a 6-point Likert scale (Jingushi, 2024b). Unfortunately, Participant D could not record well during this survey. Figures 2 and 3, originally developed in Jingushi (2024b), illustrate the participants' perspectives on generative AI and their levels of satisfaction with each prompt.

Figure-1. The number of words of each prompt

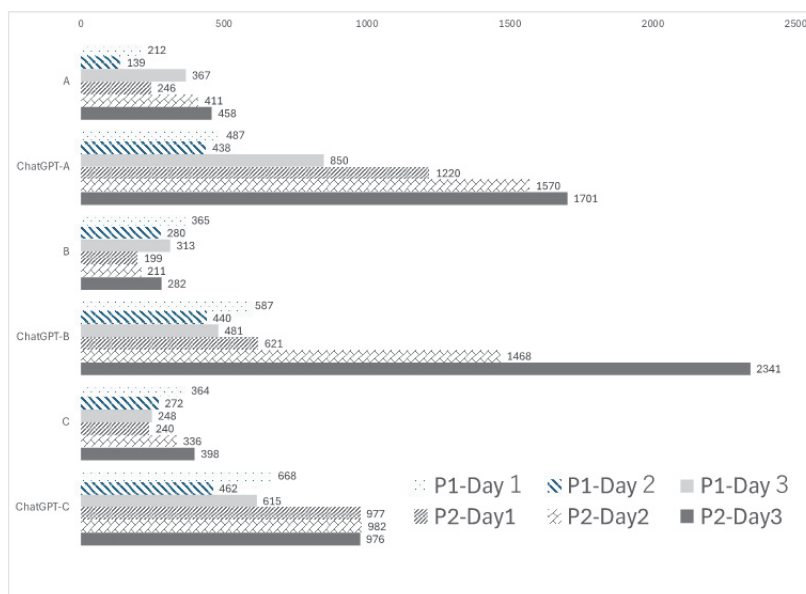


Figure-2. Web Survey Results

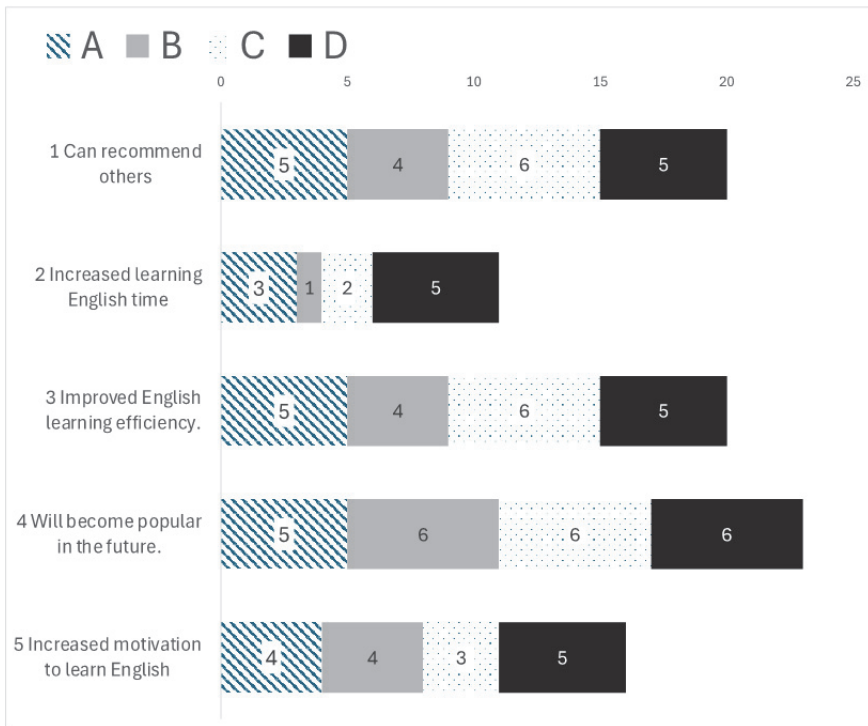
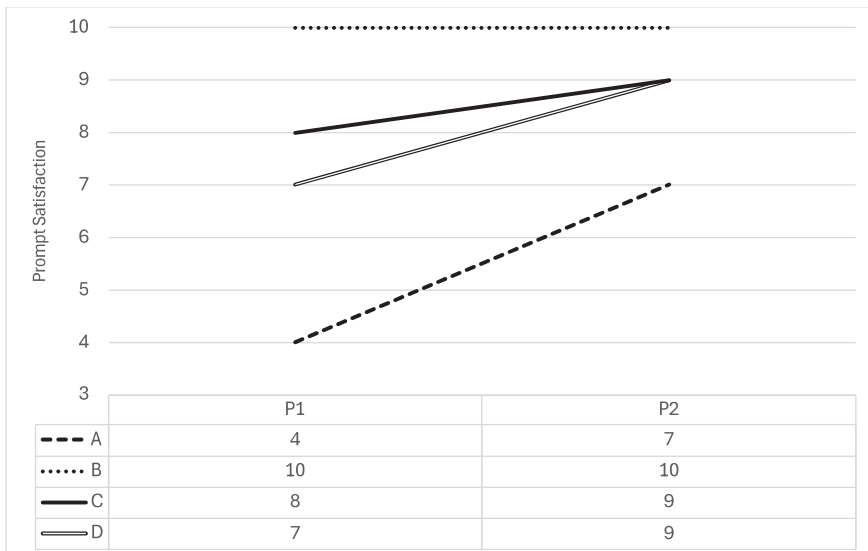


Figure-3. Prompt Satisfaction



8. Discussion

This study aimed to explore two key aspects of using generative AI tools for English speaking practice: RQ1 and 2.

Regarding RQ1 : Is there a difference in the number of words spoken by learners when using two different types of prompts?, the results from three participants (A, B, and C) demonstrated that the type and design of prompts had a notable effect on the number of words generated during interactions with ChatGPT. For example, Participant A, an advanced English speaker (CEFR C1), produced a consistently high word count in response to more abstract and complex prompts, suggesting that intellectually engaging tasks encouraged extended output.

Participant B, at the upper-intermediate level (CEFR B2), showed fluctuation in word count depending on the difficulty of the prompt. Prompts that matched his proficiency level led to more sustained conversations, whereas overly difficult prompts seemed to hinder output. This highlights the need to design prompts that are appropriately scaffolded for intermediate learners.

Participant C, who frequently uses English at work and regularly utilizes AI tools, maintained consistent output across prompt types. His familiarity with AI tools and comfort using English likely contributed to his stable interaction with the system.

Participant D, a high school English teacher with a B2 proficiency level, encountered some technical difficulties during ChatGPT interactions, and as such, her word count data could not be reliably collected. Nevertheless, her inclusion expands the participant pool beyond Jingushi (2024b) and introduces the valuable perspective of a practicing teacher, thereby strengthening the reliability and diversity of the dataset..

Regarding RQ2: How can generative AI tools influence learners' motivation in English speaking practice?, figure 2 (Jingushi, 2024b) showed that all participants—including Participant D—reported positive motivational effects from using generative AI for speaking practice. For example:

Participant A found the tasks intellectually stimulating and aligned with personal goals, which enhanced intrinsic motivation. On the other hand, participant B was sensitive to task complexity but remained engaged when the prompts felt achievable, supporting motivation through balanced challenge. Regarding the participant C, he expressed strong practical appreciation of the tool, suggesting that real-world relevance and ease of use contribute to higher motivation. Finally, participant D, despite the issues with prompt execution, gave high scores on key indicators: His data suggests that well-constructed prompts, even when not used perfectly, can still positively shape learners' attitudes toward AI-supported language learning.

These findings align with Self-Determination Theory(Deci & Ryan, 2013), which

emphasizes the importance of autonomy, competence, and relatedness in fostering intrinsic motivation. The use of prompt engineering to tailor tasks to the learner's level and interests directly supports these three psychological needs.

As noted by Jingushi (2024a), there are more likely to be variations in motivation based on learners' various backgrounds, and each learner's use of appropriate prompts may allow for learning to be tailored to their individual personalities and characteristics. The present study supports this claim by showing that even learners with similar proficiency levels (B2), like Participants B and D, respond differently depending on their teaching context, AI familiarity, and expectations.

While the current study employed similar instruments to Jingushi (2024b), including web-based surveys and Zoom interviews, it extends the previous work in two important ways. First, the participant pool was expanded by adding Participant D, a high school English teacher, thereby introducing a new perspective from the educational field and strengthening the reliability of the findings. Second, the re-analysis of both the original and new data revealed novel insights: unlike the highly motivated learners in Jingushi (2024b), Participant D reported increased study time, higher motivation, and a willingness to recommend generative AI to others. These findings highlight that generative AI tools can positively influence learners not only at advanced proficiency levels but also at the intermediate level, offering practical implications for both learners and educators.

Overall, the findings suggest that the integration of generative AI into English speaking practice offers promising opportunities—but only when prompt design is carried out thoughtfully. Educators and learners alike must consider not only the linguistic content of the prompts but also their motivational impact. When implemented with pedagogical awareness, generative AI tools may not only facilitate practice in spoken English but also support learners' long-term motivation and confidence in using the language.

9. Conclusion

This study offers important insights into how generative AI tools—such as ChatGPT—can be utilized to enhance English speaking practice for EFL learners. By focusing on prompt engineering, it has shown that well-crafted prompts tailored to learners' proficiency levels and motivational profiles can result in more meaningful and effective spoken interactions with AI.

The findings demonstrate that generative AI has the potential to act as a personalized conversation partner, providing opportunities for repeated, low-pressure speaking

practice. Moreover, the study suggests that prompts designed with pedagogical intention can support learner autonomy and engagement, which are essential components for sustained language development.

A key takeaway from this study is the importance of considering learner diversity in both English proficiency and motivational orientation. As noted by Jingushi (2024a), there are more likely to be variations in motivation based on learners' various backgrounds, and each learner's use of appropriate prompts may allow for learning to be tailored to their individual personalities and characteristics. This highlights the need for adaptable prompt strategies that reflect each learner's unique context.

However, several limitations should be acknowledged. The study involved only four participants with varying levels of English proficiency, which, while providing some comparative depth, limits the generalizability of the findings. Furthermore, the study duration was relatively short, spanning just three weeks each. Future studies could benefit from larger participant groups and extended timelines to explore how generative AI tools influence language development over time.

Additionally, this research focused primarily on prompt types and learner responses in speaking contexts. Further investigation into how different forms of generative AI affect vocabulary acquisition, fluency, and spontaneous speaking ability would enrich the field. Longitudinal studies that examine learner motivation and performance over time would also be valuable.

In summary, this study lays the groundwork for understanding how generative AI tools can be strategically integrated into EFL speaking instruction. By aligning prompt design with learner needs and motivations, educators can maximize the benefits of AI for speaking development. Continued research in this area can contribute to the creation of more personalized, engaging, and effective English speaking environments that reflect the evolving demands of global communication.

10. References

- Ali, F., Choy, D., Divaharan, S., Tay, H. Y., & Chen, W. (2023). Supporting self-directed learning and self-assessment using TeacherGAIA, a generative AI chatbot application: Learning approaches and prompt engineering. *Learning: Research and Practice*, 9(2), 135-147.
- Brown, T. B., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877- 1901.
<https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a>

-Paper.pdf

- Deci, E. L., & Ryan, R. M. (2013). Intrinsic motivation and self-determination in human behavior. *Springer Science & Business Media*.
- Han, J., Yoo, H., Myung, J., Kim, M., Lee, T. Y., Ahn, S. Y., ... & Answer, A. N. (2023). Exploring student-ChatGPT dialogue in EFL writing education. In *Thirty-seventh conference on neural information processing systems, neural information processing systems foundation*.
- Jingushi, K. (2024a). Motivational Change Collaborated by Ideal L2 Self and Self-Regulated Learning: Comprehensive Observation Based on the Complex Dynamic System Theory. *Language, Communication, and Culture*. 21(1), 65-78.
- Jingushi, K. (2024b). Prompting Relationship: Using AI for EFL Motivation. *GEN TEFL Journal Vol. 9*
- Josh F. (2025). Student Generative AI Survey 2025. *HEPI Policy Note 61*.
- Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 30, 121–204.
<https://doi.org/10.1007/s40593-019-00186-y>
- Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2024). Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in english education. *Education and Information Technologies*, 29(9), 11483-11515.
- Li, Y. (2023). A practical survey on zero-shot prompt design for in-context learning. *arXiv preprint arXiv:2309.13205*.
- Liu, X., et al. (2021). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1-35.
<https://arxiv.org/abs/2107.13586>
- Nitta, R., & Baba, K. (2015). Self-regulation in the evolution of the Ideal L2 Self: A complex dynamic systems approach to the L2 motivational self system. *Motivational dynamics in language learning*, 367-396.
- Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (pp. 1-7).
- Shishido, J. C. Y. M. (2024). Prompting Brilliance: Unlocking ChatGPT's Potential to Revolutionize EFL Dialogue Practices. *The Association for Natural Language Processing, Proceedings of the 30th Annual Meeting*. 3155-3159.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35, 24824-24837.
- Young, J. C., & Shishido, M. (2023). Investigating OpenAI's ChatGPT potentials in generating Chatbot's dialogue for English as a foreign language learning.

International Journal of Advanced Computer Science and Applications, 14(6).

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, 2021(1), 8812542. <https://doi.org/10.1155/2021/8812542>

Zheng, Y., Wang, Y., Liu, K. S. X., & Jiang, M. Y. C. (2024). Examining the moderating effect of motivation on technology acceptance of generative AI for English as a foreign language learning. *Education and Information Technologies*, 29(17), 23547-23575. <https://doi.org/10.1007/s10639-024-12763-3>

11. Appendices

All of materials, including results of the survey and script of the conversation with Chat GPT are stored on the following link.

https://drive.google.com/drive/folders/1mCsHcXzJZgnD0wIXnjd0-Yijooyoy3?usp=drive_link

