INTERNATIONAL CALL RESEARCH CONFERENCE

https://doi.org/10.29140/9780648184485-13

LinguaPilot: A GAI-Based Framework for Self-Regulated EFL Reading

Baorong Huang*
Institute of Corpus Studies and Applications, Shanghai International Studies University, China

Abstract

Generative artificial intelligence (AI) shows great promise for computer-assisted language learning (CALL), but its native application may lead to limited benefits. This study introduces LinguaPilot, an innovative framework and online platform that integrates generative AI technologies to enhance self-regulated learning (SRL) among university-level English as a Foreign Language (EFL) learners. Following Zimmerman's Cyclical Phases Model, LinguaPilot incorporates AI features into pre-reading, during-reading, and post-reading stages, corresponding to forethought, performance, and self-reflection phases, respectively. In addition, as personalized recommendations and quizzes are two essential elements in self-regulated learning, this study experimented with various optimization techniques and reported the best results, demonstrating effective methods for harnessing AI's potential to improve self-regulated EFL reading. Accessible via an intuitive online interface, LinguaPilot makes a tentative fulfillment of AI's transformative role in enhancing self-regulated language learning and the application of generative AI in EFL reading.

Keywords: Generative artificial intelligence, EFL reading framework, Self-regulated learning

Introduction

Generative artificial intelligence (GAI) technologies release huge potential in computer-assisted language learning. However, the direct application of ChatGPT and similar generative AI technologies in language learning, without

*Corresponding Author E-mail: 0214101730@shisu.edu.cn

In Y. Wang & M. S. Cárdenas Claros (Eds.), Proceedings of the XXIInd International CALL Research Conference, 2024 (pp. 85–90). Castledown.

proper monitoring and control, will produce limited effects. Against this background, this study proposes an innovative framework and an online platform named LinguaPilot, which incorporates generative AI technologies to guide English as a Foreign Language (EFL) learners at the university level in fostering self-regulated learning capabilities.

Aligned with the Cyclical Phases Model proposed by Zimmerman (1996), the framework integrates the generative AI features into the three integral processes of reading—namely, pre-reading, during-reading, and post-reading—which correspond with the forethought, performance, and self-reflection phases, respectively. In addition, inspired by the roles of AI interventions in stimulating self-regulated learning (SRL) processes proposed by Järvelä et al. (2023), we identified three roles in AI-assisted self-regulated reading (assistants, tutors, and peer learners) and implemented them in the framework. Prompt engineering, fine-tuning, and multiple-agent techniques for text complexity classification and question generations are compared to gain the best methods, and the practical implications for leveraging large language models (LLMs) in self-regulated EFL reading are also discussed.

Description of the GAI-based Framework for Self-regulated Reading

Reading is an essential activity in language learning, but it isn't easy to use LLMs like ChatGPT to cover the whole gamut of the reading. LinguaPilot, a dedicated web-based online reading platform based on the theoretical framework of self-regulated learning, has been developed to guide and monitor the reading activities of EFL learners, as shown in Figure 1.

This framework provides comprehensive support for self-regulated learning and is designed based on the previous empirical studies that demonstrated the potential of advanced learning technologies for supporting self-regulated learning, including learning analytics dashboard, goal-setting support, self-assessment activities, initiating reflection, and personalized recommendations as categorized by Radović et al. (2024) more efficient learning progress, and other non-academic benefits than those who do not. However, students often have poor self-regulation practices, lack reflective thinking, and fail to monitor their learning against defined goals. Therefore, a variety of advanced learning technologies and features have been developed to support students with regulation. Nevertheless, the optimal level of regulation support and the most effective combination of these features remain unestablished in the existing research literature. Therefore, this study employs quantitative research to examine the effectiveness of two learning environments with different levels of self-regulation support (less and more. In particular, personalized recommendation is achieved

Figure 1. Architecture of LinguaPilot.

partially via an optimized text complexity classification algorithm and quizzes in the self-assessment activities are generated by GAI.

In addition, in the framework, a chatbot backed by GPT3.5 (OpenAI, 2023) from OpenAI API is used to assist learners when they have doubts about reading materials, help them analyze sentence structures, and translate difficult sentences into their native language. Apart from using the system prompts to define the role of LLMs as assistants, tutors, and peer learners, the framework optimized the text complexity classification algorithm and question generation for self-assessment through multi-agent architecture (Durante et al., 2024).

As stated in the input hypothesis (Krashen, 1977), learners improve their language skills by understanding the slightly more complex input than their current proficiency level. The majority of EFL learners in Chinese universities need to sit for CET4 (College English Test Band 4), CET6 or TEM4 (Test for English Majors, Band 4) and TEM8 exams. To recommend suitable reading materials, we applied fine-tuning on deep learning models with a dataset of 1320 articles, consisting of articles from authentic exam materials and study guides. Then, we trained the deep-learning models with that dataset. Through extensive experiments, we found that the model fine-tuned with RoBERTa-large (Liu et al., 2019) achieved the highest accuracy at 81% with three difficulty levels (CET4, CET6, and CET8), better than the supported vector machine models, random forest models, and XGBoost models trained with the features based on the syntactic complexity features (Lu, 2011, 2017) or the comprehensive features extracted by Lingfeat (Lee et al., 2021). In addition,

large language models, such as Llama3 (Touvron et al., 2023), an open-source LLM released by Meta, or ErnieBot 3.5-8K (Sun et al., 2019), a closed-source LLM released by Baidu, performed rather poorly in this difficulty classification task, with the accuracy below 50%.

Good self-assessment questions are important for learners to monitor their understanding of reading materials. To improve the quality of the generated questions, the framework proposed a multi-agent method for question generation. In the multi-agent method, one question writer agent and one question reviewer agent were used, in which the question reviewer agent would review the questions and give suggestions or criticism for the question writer agent to rewrite the question distractors. We have conducted experiments with Qwen (Bai et al., 2023), a large language model (LLM) released by Alibaba, including its variants qwen_plus and qwen_max. Inspired by the practice of using LLM for judging LLM (Zheng et al., 2023), we scored the generated results by using GPT4o, a powerful LLM released by OpenAI. Twenty articles from CET4 and CET6, 10 for each, were randomly sampled from the exam dataset. Then, we prompted the two LLM models to generate five questions for each article and also four options, including three distractors for each question. The quality of questions was evaluated on a scale from 1 to 10 in three dimensions: Fluency,

Table 1. Comparison of question generation results

Model	Total Questions	Fluency Mean(SD)	Semantics Mean(SD)	Relevance Mean(SD)	Plausibility Mean(SD)	Reliability Mean(SD)
qwen_max + guidelines + agent	86	–	–	–	2.80(1.09)	2.76(1.05)
qwen_ max + guidelines	75	10(0)	9.53(0.50)	9.99(0.12)	2.65(0.94)	2.63(0.93)
qwen_ max + agent	92	–	–	–	2.91(1.11)	2.86(1.09)
qwen_max	80	10(0)	9.60(0.52)	9.91(0.48)	2.73(0.95)	2.70(0.93)
qwen_plus + guidelines + agent	100	–	–	–	3.06(1.22)	2.98(1.19)
qwen_plus + guidelines	92	10(0)	9.49(0.50)	9.95(0.23)	2.97(1.33)	2.89(1.32)
qwen_plus + agent	95	–	–	–	2.85(1.20)	2.78(1.18)
qwen_plus	87	10(0)	9.47(0.52)	9.92(0.41)	2.72(1.21)	2.68(1.16)

Semantics, and Relevance, as proposed by Zou et al. (2022). In addition, the quality of the distractors was scored on a scale from 1 to 5 against plausibility and reliability (Haladyna et al., 2002; Pho et al., 2014). As shown in Table 1, the two LLM variants, qwen_plus and Q qwen_max, failed to generate the exact number of questions (100 in total) for the 20 articles. Although feedback from the reviewer agent led to an increase in the total number of questions, they still fell short of the required count. For distractor generation, all models with reviewer agents performed better than their counterparts, i.e., corresponding no-agent models, which demonstrated the effectiveness of the agent method.

Conclusion

This study demonstrates the pivotal role of generative AI technology in assisting EFL learners with reading, particularly within the context of learner autonomy. Based on the theoretical framework of self-regulated learning, this study weaved the principles of theory into the design and implementation of an online reading platform, with considerable efforts in algorithm optimizations to meet the specific needs of Chinese EFL learners. However, this study focuses on the technical details of the implementation of LLM as a pilot in self-regulated reading. In future, empirical studies should be conducted to ascertain the effectiveness of this framework in real-world scenarios.

References

Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F., Hui, B., Ji, L., Li, M., Lin, J., Lin, R., Liu, D., Liu, G., Lu, C., Lu, K., ... Zhu, T. (2023). Qwen technical report. *arXiv*. http://arxiv.org/abs/2309.16609

Durante, Z., Huang, Q., Wake, N., Gong, R., Park, J. S., Sarkar, B., Taori, R., Noda, Y., Terzopoulos, D., Choi, Y., Ikeuchi, K., Vo, H., Fei-Fei, L., & Gao, J. (2024). Agent AI: surveying the horizons of multimodal interaction. *arXiv*. http://arxiv.org/abs/2401.03568

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309–334. https://doi.org/10.1207/S15324818AME1503_5

Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, *54*(5), 1057–1076. https://doi.org/10.1111/bjet.13325

Krashen, S. (1977). Some issues relating to the monitor model. In Brown, H., Yorio, C. and Crymes, R. (eds.). *Teaching and learning English as a second language: some trends in research and practice*. Washington, DC: TESOL, 144–48.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv*. http://arxiv.org/abs/1907.11692

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36–62. https://doi.org/10.5054/tq.2011.240859

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, *34*(4), 493–511. https://doi.org/10.1177/0265532217710675

OpenAI. (2023). GPT-4 technical report (arXiv:2303.08774). *arXiv*. http://arxiv.org/abs/2303.08774

Pho, V.-M., Andre, T., Ligozat, A.-L., Grau, B., Illouz, G., & Francois, T. (2014). Multiple choice question corpus analysis for distractor characterization. *Proceedings of Ninth International Conference on Language Resources and Evaluation*. LREC2014.

Radović, S., Seidel, N., Menze, D., & Kasakowskij, R. (2024). Investigating the effects of different levels of students' regulation support on learning process and outcome: In search of the optimal level of support for self-regulated learning. *Computers & Education, 215*, 105041. https://doi.org/10.1016/j.compedu.2024.105041

Sun, Y., Wang, S., Li, Y., Feng, S., Chen, X., Zhang, H., Tian, X., Zhu, D., Tian, H., & Wu, H. (2019). ERNIE: Enhanced Representation through Knowledge Integration. *arXiv*. http://arxiv.org/abs/1904.09223

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *arXiv*. http://arxiv.org/abs/2306.05685

Zou, B., Li, P., Pan, L., & Aw, A. T. (2022). Automatic true/false question generation for educational prpose. *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, 61–70. https://doi.org/10.18653/v1/2022.bea-1.10

INTERNATIONAL CALL RESEARCH CONFERENCE

https://doi.org/10.29140/9780648184485-14

Online Teacher Communities in the Post-Pandemic Era: How ChatGPT-Focused Online Teacher Communities are Assisting Language Teachers

Yurika Ito*
Kanagawa University, Japan

Abstract

To keep up with the changes brought by technological advances and surrounding social situations, many language teachers around the world have been turning to online teacher communities formed on various social media platforms, including Facebook, as a source of professional and emotional support. In particular, the popularity of online teacher communities heightened during the COVID-19 pandemic when many language teachers around the world were forced to temporarily shift to remote teaching. While the pandemic has waned and in-person classes have largely resumed, many teachers are now grappling with a new challenge: Understanding the role of ChatGPT and other generative AI technologies in the language learning classroom. As such, teachers in the post-pandemic era appear to be joining newly established ChatGPT-focused online teacher communities to explore how generative AI technologies can be utilized for educational purposes. The main objective of this ongoing research is therefore to understand how language teachers can capitalize upon self-generated ChatGPT-focused online teacher communities as a means of learning about how to use ChatGPT in educational settings. Employing a mixed-methods research design, data were collected via online observations of a self-generated Facebook group consisting of teachers who were interested in learning about ChatGPT and self-reported instruments, including a questionnaire and in-depth interviews with language teachers. The findings are discussed in terms of the potential benefits and drawbacks of leveraging a ChatGPT-focused online teacher community on Facebook as a way of

*Corresponding Author E-mail: yurika.ito@akane.waseda.jp

In Y. Wang & M. S. Cárdenas Claros (Eds.), Proceedings of the XXIInd International CALL Research Conference, 2024 (pp. 91–96). Castledown.