Intelligent Error Management Empowered by Large Language Model-based Agent in EFL Education of Junior High

Jiahui Hou and Yingchun Ren*

China University of Petroleum (East China), Qingdao, Shandong, 266580, China

*yingchunren@126.com

Abstract. This study examines the implementation pathways of large language model (LLM)-based agent for English error management in EFL education, comparing two LLM-based agent systems: a plugin-augmented agent (EEM_PA) and a structured workflow-driven agent (EEM_WF). Both systems were developed using the DeepSeek model and deployed on the low-code agent configuration platform COZE. Through a comparative evaluation focused on functional performance and user experience among ten Chinese junior high students with similar English proficiency, the study found that while both systems demonstrated comparable functional effectiveness, the plugin-augmented EEM_PA outperformed the workflow-driven EEM_WF in terms of user experience. This suggests that the workflow-driven design, though systematic and potentially useful for standardized review, may function more as a supportive tool for teachers or parents rather than as an engaging learning companion for students. This study highlights practical pathways for educators to develop intelligent tutoring systems by leveraging LLMs and low-code platforms. It offers a comparative reference for designing AI-assisted learning tools that balance interactive engagement with structured pedagogical support.

Keywords: LLM-based agent; intelligent English error management; plugin-augmented; workflow-driven

1 Introduction

LLM-based agents, an upgraded AI technology, have emerged as dynamic systems that maintain state across multiple LLM calls and integrate external tools, enabling multi-step workflows and consistent interactions [28], displaying great potential in English as Foreign Language (EFL) educational [27]. Furthermore, low-code agent development platforms (like COZE) have significantly reduced the barriers to access LLMs and agent techniques, enabling educators to easily create AI assistants capable of supporting teaching and learning process [15]. Meanwhile, personalized learning has emerged as a critical approach in EFL education to address student diversity [10]. A

S. Garcia-Esteban et al. (eds.), *Proceedings of the 2026 5th International Conference on Social Sciences and Humanities and Arts (SSHA 2026)*, Advances in Social Science, Education and Humanities Research 1014,
https://doi.org/10.2991/978-2-38476-577-5_48

key yet often overlooked arena for its implementation is error management—a micro-level practice essential for foundational knowledge acquisition.

This paper explores the integration of LLM-based agents into this domain. By developing and evaluating two self-designed intelligent English error management systems EEM_PA and EEM_WF, this study aims to provide educators with actional insights into how LLM-based agents can be integrated into English error management, and the differences in the aspects of functional performance and user experience of distinct technical pathways. The following research questions guide this exploration:

RQ 1: How to design and develop intelligent English error management systems EEM_PA and EEM_WF on COZE?

RQ 2: What are the functional performance and user experience of the two English error management systems?

RQ 3: What are the differences between EEM_PA and EEM_WF from the learner's perspective?

2 Literature Review

LLM-based agents combine the strengths of LLMs with external tools and resources to enable more dynamic and autonomous operations [12]. Such agents are able to orchestrate complex educational workflows while enhancing understanding, engagement, and personalization, thereby enabling more adaptive, efficient, and scalable learning. Many existing research has introduced LLM-based agents in education field. Chu et al [4] categorized educational agents into two broad class: pedagogical agents and domain-specific educational agents, and provided a comprehensive review of them. In terms of EFL education, LLM-based agents are generally employed in reading [27], writing [9], speaking [19], and translation [6], etc.

LLM-based agents are designed to tackle complex problems, but this does not imply that they are inherently complex or difficult to develop. With the advent of low-code/no-code agent development platforms, such as COZE, Dify, and MetaGPT, individuals can now design and deploy custom agents for personal or professional tasks through intuitive visual interfaces, without the need for deep programming expertise. Certainly, the LLM that incorporated into the agent taking effect relies on precise and detailed prompts, while the whole architecture process complete also with the help of external tools.

Error management, a personalized learning activity rooted in knowledge management [7], emerges as an ideal scenario for implementing AI-driven micro-level teaching practices. Engaging with errors promotes deeper cognitive processing and exploratory active learning [16]. Empirical studies have confirmed that generating errors followed by corrective feedback strengthens memory for correct responses [20]. At the junior high school level, cultivating the habit of analyzing and systematizing errors from daily exercises and tests is crucial for consolidating foundational knowledge and developing effective learning strategies. Researchers and educators have explored AI-assisted error analysis and remediation through smart mistake notebooks [29], homework helpers [5], and AI-enhanced error logging and note-taking processes [1, 3].

Nevertheless, significant research gaps remain. First, the value of error management is mostly stressed on science subjects like math [23] while the discussions on English error management are still in a small number. Second, although LLM-based agents have been widely used across various EFL domains [19], comparative studies examining different AI technical approaches, such as a plugin-augmented agent and a workflow-driven agent, within the specific context of error management are notably scarce. This study aims to conduct a comparative investigation into the application of LLM-based agents in English error management at the junior high school level, with the goal of elucidating their practical efficacy in this foundational yet critical educational domain.

3 Methods

3.1 Design and Development of English Error Management System EEM_PA and EEM_WF

This study designed two LLM-based agents to display different technical pathways for intelligent English error management on the COZE platform, a low-code agent development environment where building an agent primarily consists of three key modules: persona and prompts, model configuration (including model selection, skills, knowledge, memory, and dialogue), and preview. DeepSeek model is selected because of its outstanding high performance, cost-effectiveness, and strong multilingual and real-time interaction capabilities [18]. Both agents were designed to recognize the error pictures of users and fulfill three core educational functions: error diagnosis, targeted exercise generation, and structured collection of errors for subsequent review. These shared objectives guided the selection of plugins and the construction of structured workflows respectively, enabling each agent to implement the same pedagogical process through architecturally distinct means.

The Architecture of EEM_PA

Persona and prompts: 1) *Role*: The agent is designated as a senior teacher with many years of experience in junior high school English education, especially in ninth grade, detailing its teaching expertise, knowledge of common student errors in English, and a patient, friendly teaching style. This aims to guide the model to produce more authoritative and pedagogically targeted content. 2) *Skills*: The agent is explicitly given three core skills, including error diagnosis (error identification, cause analysis, mapping knowledge domain), personalized exercise generation (multi-level and multi-format exercises), learning management (error categorization, error document generation). This ensures the comprehensiveness of the system's functions. The invocation of plugins is also required. 3) *Restrictions*: Teaching content (only English), knowledge level (junior high curriculum), interaction guidelines (friendly and encouraging style, avoidance of cognitive overload), and ethical appropriateness (avoidance of complex political or religious examples) are required.

Model Configuration: 1) *Core LLM selection*: DeepSeek-V3. 2) *Plugins*: The plugin "qwen_vl_max" (Qwen image recognition) is selected as a stable and reliable solution for converting user-uploaded error images into text. To achieve persistent error management, the plugin "execute" (Dameng database) is integrated to invoke the database and achieve the storage of user's error records and learning data. The plugin "create_docx" is added for available off-line error documents for users. 3) *Memory*: a database "mistake_data1" is created. 4) *Knowledge*: A series of texts on ninth-grade fundamental knowledge and grammar were incorporated into the knowledge base to ensure the agent retrieves information from it during operation. 5) *Opening statement*: Opening prompts are set in to guide users on how to interact with the agent.

The Architecture of EEM_WF

Persona and prompts: 1) *Role*: An "error management assistant", emphasizing its task-oriented nature and the function of document generation. 2) *Skills*: There is no complex description but simple requirements like invoking the "English_notebook" workflow to process user's input. To compensate for the few seconds required for workflow execution, the agent is asked to first provide a brief preliminary analysis and reassuring message immediately after invoking the workflow, enhancing the interaction experience during the wait. The format of reply is also required.

Model Configuration: 1) *Core LLM selection*: DeepSeek-V3. 2) *Workflow design*: A customized workflow named "English_notebook" is designed, decomposing the error-processing pipeline into a series of sequential nodes, including error input -- model analysis -- report organization -- document generation, achieving full process automation. Two documents will be offered: One with exercises and the other with answers. 3) *Knowledge*: A series of texts on ninth-grade fundamental knowledge and grammar were incorporated into the knowledge base to ensure the agent retrieves information from it during operation. 4) *Opening statement*: Opening prompts are also set in to guide users on how to use the agent.

3.2 Participants and System Evaluation

To test the performance of these two LLM-based agents, ten ninth-grade students (5 male, 5 female) of similar intermediate English proficiency (average score between 72-96 in the latest test) from the same class were chosen to use the systems. Informed consent was obtained from all participants and their guardians prior to the study.

Participants used each assigned intelligent system for a week (7 days), processing 2--3 self-encountered English errors per day. At the end of each week, participants completed the functional performance and user experience questionnaires specific to that system. After experiencing both systems, a final semi-structured interview was conducted with each participant to gather their comparative perceptions and overall preferences.

The evaluation of functional performance includes three indicators: (i) Comprehensibility: Were the system's explanations and analysis clear and easy to understand? (ii) Helpfulness: Are the exercises generated by the system helpful for reinforcing the knowledge? (iii) Completeness: Did the system provide a complete procedure (diagnosis, explanation, practice) for the error analysis?

The user experience will be assessed according to System Usability Scale [25], yielding a score from 0 to 100, supplemented by interview data on user preferences.

A unified questionnaire (15 items) was used (see Appendix 2), with a 5-point Likert scale [17] for functional performance items (3-5) and SUS items (6-15) for usability.

Descriptive statistics were computed and differences between the two systems were assessed using Wilcoxon signed-rank tests [22].

4 Results

4.1 Functional Performance

Functional performance was assessed across three dimensions: comprehensibility, helpfulness and completeness. As shown in Table 1, the plugin-augmented agent (EEM_PA) received higher mean scores than the workflow-driven agent (EEM_WF) on all three indicators.

Table 1. Descriptive Statistics for Functional Performance Indicators of EEM_PA and EEM_WF

Functional Performance				
System	N	Comprehensibility (M/SD)	Helpfulness (M/SD)	Completeness (M/SD)
EEM_PA	10	4.50 (0.71)	4.70 (0.48)	4.40 (0.52)
EEM_WF	10	4.00 (0.47)	4.40 (0.52)	4.10 (0.57)

A composite functional performance score was derived by summing the three indicators. The plugin-augmented agent (EEM_PA) achieved a higher total score (M = 13.60, SD = 1.51) compared to the workflow-driven agent (EEM_WF) (M = 12.50, SD = 0.85).

Wilcoxon signed-rank tests were conducted to compare the two systems on each functional performance indicator as well as the composite score. No statistically significant differences were found between the two agents in comprehensibility ($Z = -1.667$, $p = 0.096$), helpfulness ($Z = -1.732$, $p = .083$), or completeness ($Z = -1.342$, $p = 0.180$). Similarly, the composite functional performance score showed no significant difference ($Z = -1.572$, $p = 0.116$).

4.2 User Experience

User experience was evaluated using the SUS, with scores ranging from 0 to 100. As presented in Table 2, the plugin-augmented agent (EEM_PA) obtained a higher mean SUS score (M = 79.50, SD = 12.63) compared to the workflow-driven agent (EEM_WF) (M = 68.00, SD = 14.18).

Table 2. Descriptive Statistics for SUS Scores of EEM_PA and EEM_WF

User Experience				
System	N	M/SD	Min	Max
EEM_PA	10	79.50 (12.63)	62.50	100.00
EEM_WF	10	68.00 (14.18)	35.00	87.50

A Wilcoxon signed-rank test was conducted to compare the SUS scores between the two systems. The test revealed a statistically significant difference ($Z = -2.077$, $p = 0.038$), indicating that the plugin-augmented agent (EEM_PA) provided a significantly better user experience than the workflow-driven agent (EEM_WF).

5 Discussion

5.1 Main Findings

(1) *Comparable Functional Performance Between Architectures*: Both the plugin-augmented and workflow-driven LLM-based agents demonstrated statistically similar effectiveness in fulfilling core error management functions, namely error diagnosis, targeted exercise generation, and structured error archiving. This indicates that despite differences in technical implementation, both approaches can reliably support pedagogical functions within EFL learning.

(2) *Significantly Superior User Experience of the Plugin-Augmented Agent*: The plugin-augmented agent (EEM_PA) received significantly higher SUS scores than the workflow-driven agent (EEM_WF). This suggests that although the workflow-driven agent provides systematic and template-based outputs suitable for structured review, an interactive, conversationally flexible design better aligns with students' preferences and engagement patterns.

(3) *Low-Code Platforms as Accessible Pathways for Educators*: The results demonstrate that educators can utilize low-code platforms such as COZE to design and deploy distinct types of LLM-based tutoring agents, offering a practical, low-threshold pathway for integrating AI into EFL instruction without requiring advanced programming skills.

5.2 Interpretation of the Findings

As the plugin-augmented agent gains significant advantages over the workflow-driven agent in the aspect of user experience, the disparity can be understood through the lens of student interaction patterns beyond operational problems. The plugin-augmented agent (EEM_PA), functioning as a generative chatbot, fostered a more dynamic and engaging dialogue, which appeared to stimulate greater learning interest and participation among students [8, 14, 26,], creating meaningful learning experiences [13]. In contrast, interactions with the workflow-driven agent (EEM_WF) were markedly diminished. Although it supported conversation, students predominantly terminated the interaction immediately after downloading the generated document. The entire process was characterized by a passive wait for the output, lacking a sense of involvement and failing to motivate further interaction. However, motivation and engagement are vital because they will make students better cope with the difficulties and then improve their academic performance [21, 24]. Though an agent with a structured workflow can improve processing efficiency in complex scenarios and contribute to better educational outcomes [11], it suggests that such a streamlined mode might not be good as a direct learning companion for juvenile students. Instead, it may be better suitable for use by parents or teachers so that they can easily acquire useful and personalized learning materials. Afterall, those adults can maintain a balance between the collaboration with intelligent tutoring tool and individual learning activities [2].

6 Conclusion

This study designed, implemented, and evaluated two LLM-based English error management agents, one plugin-augmented (EEM_PA) and one workflow-driven (EEM_WF), using the low-code platform COZE. Both agents were developed with the same underlying model (DeepSeek) and were designed to fulfill three core pedagogical functions: error diagnosis, targeted exercise generation, and structured error archiving. Results showed that while the two systems achieved comparable functional performance, the plugin-augmented agent received significantly higher usability ratings on the SUS, indicating a clear preference among students for its interactive and conversational style.

The findings offer practical insights for educators and AI tool designers. First, they demonstrate that educators can effectively build functional AI-assisted learning tools using low-code platforms without advanced programming expertise. Second, plugin-augmented agents better support student engagement through flexible dialogue while the workflow-driven agents can deliver structured and consistent outputs may be suitable for teacher- or parent-led review. This suggests that the choice of agent design should align with the intended user and instructional context.

This study has several limitations. First, the sample size was small and confined to junior high students from a single region, which may limit generalizability. Second, the evaluation period was brief, and no longitudinal data on learning outcomes were collected. Finally, reliance on subjective self-reported measures precluded the inclusion of objective performance metrics. Future research could expand participant diversity, employ longitudinal designs with objective learning indicators, and explore hybrid architectures that integrate workflow structure with interactive plugin functionality.

Reference

1. Alisoy, H. : Can NotebookLM Support English Language Learners? A Theoretical Perspective on AI Tools in Education. *Porta Universorum, 1*(6), 25-55 (2025). https://doi.org/10.69760/portuni.0106003.
2. Armamento, P. J., Huynh, K. A., & Rohan, A. : Empowering tech education: harnessing prompt engineering techniques in large language models. *Proceedings of the 17th Annual International Conference of Education, Research and Innovation*. 8817-8826 (2024). doi: 10.21125/iceri.2024.2217.
3. Chen, M. R. A. : The AI chatbot interaction for semantic learning: A collaborative note-taking approach with EFL students. *Language Learning & Technology, 28*(1), 1-25 (2024). https://doi.org/10.64152/10125/73586.
4. Chu, Z. D., Wang, S., Xie, J., Zhu, T. H., Yan, Y. B., Ye, J. H., et al. : Llm agents for education: Advances and applications. arXiv preprint arXiv:2503.11733 (2025). https://doi.org/10.48550/arXiv.2503.11733.

5. Davis, S. R., DeCapito, C., Nelson, E., Sharma, K., & Hand, E. M. : Homework helper: Providing valuable feedback on math mistakes. *Proceedings of the International Symposium on Visual Computing (ISVC)*. 533-544 (2020). https://doi.org/10.1007/978-3-030-64559-5_42.
6. Guo, S. T., Zhang, S. L., Ma, Z. R., Zhang, M., & Feng, Y. : Agent-simt: Agent-assisted simultaneous machine translation with large language models. arXiv preprint arXiv:2406.06910 (2024). https://doi.org/10.48550/arXiv.2406.06910.
7. Hargreaves, D. H. : The knowledge-creating school. *British journal of educational studies, 47*(2), 122-144 (1999). https://doi.org/10.1111/1467-8527.00107.
8. Hu, W. Q., Gong, R. S., Wu, S. S., & Li, Y. Y. : A conversational agent based on contingent teaching model to support collaborative learning activities: impacts on students' learning performance, self-efficacy and perceptions. *Educational technology research and development* (2025). https://doi.org/10.1007/s11423-025-10526-6.
9. Hu, Z., Chan, H. P., Li, J., & Yin, Y. : Debate-to-write: A persona-driven multi-agent framework for diverse argument generation. In *Proceedings of the 31st International Conference on Computational Linguistics*, 4689-4703 (2024). https://aclanthology.org/2025.coling-main.314/.
10. Hussain, S. M., Satti, S. M. J., & Khan, Z. : AI-Powered Personalized Learning: Advancing Language Education in the Digital Era. *Journal of Social Signs Review, 2*(4), 730-740 (2024). Retrieved from https://socialsignsreivew.com/index.php/12/article/view/71.
11. Jiang, Y. H., Lu, Y. J., Dai, L., Wang, J. T., Li, R. J., & Jiang, B. Agentic workflow for education: concepts and applications. *arXiv preprint arXiv:2509.01517* (2025). https://doi.org/10.48550/arXiv.2509.01517.
12. Jin, H. L., Huang, L. H., Cai, H. P., Yan, J., Li, B., & Chen, H. M. : From llms to llm-based agents for software engineering: A survey of current, challenges and future. *arXiv preprint arXiv:2408.02479* (2024). https://doi.org/10.48550/arXiv.2408.02479.
13. Kamalov, F., Calonge, D. S., Smail, L., Azizov, D., Thadani, D. R., Kwong, T., et al. : Evolution of AI in education: agentic workflows. *arXiv preprint arXiv:2504.20082* (2025). https://doi.org/10.48550/arXiv.2504.20082.
14. Li, H., & Graesser, A. : Impact of pedagogical agents' conversational formality on learning and engagement. *Proceedings of the International Conference on Artificial Intelligence in Education (AIED)*. 188-200 (2017). https://doi.org/10.1007/978-3-319-61425-0_16.
15. Liao, H. : A practical path of multi-AI agents collaboration based on low-code coze platform to empower students' English learning efficiency. *World Journal of Innovation and Modern Technology (WJIMT), 8*(7), 59-67 (2025). https://doi.org/10.53469/wjimt.2025.08 (07). 11.
16. Metcalfe, J. : Learning from errors. *Annual review of psychology, 68*(1), 465-489 (2017).
17. Mirahmadizadeh, A., Delam, H., Seif, M., & Bahrami, R. : Designing, constructing, and analyzing Likert scale data. *Journal of Education and Community Health, 5*(3), 63-72 (2018). doi:10.21859/jech.5.3.63.
18. Mohammed, A. A., Mudhsh, B. A., Bin-Hady, W. R. A., & Al-Tamimi, A. S. : DeepSeek and Grok in the spotlight after ChatGPT in English education: A review study. *Journal of English Studies in Arabia Felix, 4*(1), 13-22 (2025). https://doi.org/10.56540/jesaf.v4i1.114
19. Pan, M. X., Kitson, A., Wan, H. Y., & Prpa, M. : Ellma-t: an embodied llm-agent for supporting english language learning in social VR. *Proceedings of the Proceedings of the 2025 ACM Designing Interactive Systems Conference*. 576-594 (2025). https://doi.org/10.1145/3715336.3735786.
20. Potts, R., & Shanks, D. R. : The benefit of generating errors during learning. *Journal of Experimental Psychology: General, 143*(2), 644-667 (2014).

21. Rahimi, A. R., & Sevilla-Pavón, A. : The role of interactive, constructive, active, and passive learning activities (ICAPCALL) in shaping students' online engagement, and learning approaches to virtual exchange (SAVE): a bisymmetric approach. *Smart Learning Environments, 12*(61), 1-32 (2025). https://doi.org/10.1186/s40561-024-00363-0.
22. Rosner, B., Glynn, R. J., & Lee, M. L. T. : The Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics, 62*(1), 185-192 (2006). https://doi.org/10.1111/j.1541-0420.2005.00389.x.
23. Rushton, S. J. : Teaching and learning mathematics through error analysis. *Fields Mathematics Education Journal, 3*(1), 1-12 (2018). https://doi.org/10.1186/s40928-018-0009-y
24. Talebi, M., & Safaei, M. : Measuring user engagement in educational platforms: A study on interactive vs. passive learning. *International Journal of Advanced Human Computer Interaction, 1*(1), 53-62 (2022). https://www.ijahci.com/index.php/ijahci/article/view/20.
25. Vlachogianni, P., & Tselios, N. : Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *Journal of Research on Technology in Education, 54*(3), 392-409 (2022). https://doi.org/10.1080/15391523.2020.1867938.
26. Xie, T., Liu, R., Chen, Y., & Liu, G. : MOCA: A motivational online conversational agent for improving student engagement in collaborative learning. *IEEE Transactions on Learning Technologies, 14*(5), 653-664 (2021). doi: 10.1109/TLT.2021.3129800.
27. Xiao, F. W., Li, Z. H., Lin, J. J., Zou, X. H., Yang, D. B., Zou, E. W., et al. : Leveraging an LLM-enhanced bilingual conversational agent for EFL children's dialogic reading: Insights from children, parents, and educators. *Computers and Education: Artificial Intelligence* (2025). https://doi.org/10.1016/j.caeai.2025.100484.
28. Yehudai, A., Eden, L., Li, A., Uziel, G., Zhao, Y. L., Bar-Haim, R., et al. : Survey on evaluation of llm-based agents. *arXiv preprint arXiv:2503.16416*, 1-20 (2025). https://doi.org/10.48550/arXiv.2503.16416.
29. Yu, Z., Cen, G., Zhao, L., & Zhu, C : Design of smart mistake notebook based on AI and big data. *Proceedings of the International Conference on Computer Science and Education*. 18-28 (2022). https://doi.org/10.1007/978-981-99-2443-1_2.