

Large Language Models and Their Impact on English Teaching Practices

Dedakhanov Abdumalik Mutalliyevich^{1*}

^{1*}Senior Lecturer of Philology, Faculty of Humanities & Pedagogy, Turan International University, Namangan, Uzbekistan. E-mail: abdumalikdedahanov216@gmail.com,
Orcid: <https://orcid.org/0009-0009-5050-2424>

Abstract: The rapid development of Large Language Models (LLMs) like the conversational systems based on GPT has transformed from a research curiosity to a tool that can now be found in the daily lives of English language teachers and learners, raising pressing questions about how it can be best used in teaching and learning. This paper reviews the potential for the use of the LLM capabilities (such as automated feedback generation, error detection, and conversational practice) to transform English teaching workflows, and presents results from a classroom-based evaluation of an LLM-supported teaching support system. The design was a mixed-method quasi-experimental design, with 18 English teachers using LLM-generated feedback and lesson support materials for 10 weeks, and 18 English teachers who had been using a set of lesson-planning and feedback methods for a period of time continuing to do so without the use of LLM support materials. The data were collected using pre- and post-intervention teacher surveys, learner writing samples, classroom observation (notes), and semi-structured interviews, and then analyzed using descriptive statistics and thematic coding. Results reveal that educators employing the language model-assisted system stated that preparation time declined and provided students with individual feedback more frequently, while students in their classes demonstrated a more considerable improvement in grammar and revision than the control group. The suggested model incorporates the content-generation level, the level of grammar and bias supervision, the teacher-administered system, and the classroom-based feedback cycle, putting teachers' decision-making in the center of language model utilization instead of treating the system as an independent teacher. The investigation raised certain questions regarding the high dependence on the system, reliability of the output, and changes in the professional self-conception of educators acting as the evaluators of the work done. To sum up, it can be said that large language models can enhance efficiency and personalization in English teaching if used in the framework of teacher supervision and control, while it is essential to avoid direct application of the language models and ensure ongoing learning of teachers and continuous ability to monitor the models' output.

Key Words: large language models, english language teaching, generative AI, automated feedback, teacher professional identity, classroom integration, natural language processing

(Received: 03 July 2023; Revised: 21 July 2023; Accepted: 31 August 2023; Published: 29 September 2023)

Introduction

English language learning classrooms have experienced a transformation marked by the emergence of numerous AI tools that facilitate various aspects of the English language learning process, including AI-supported lesson planning and AI-enabled simulated speech and writing interactions with students (Topsakal & Topsakal, 2022). Early AR-VI-GLM applications in the education sector demonstrate the rapid advancement of these technologies, from experimental prototypes to classroom-worthy tools designed for use with young learners, increasing the importance of grasping potential pedagogical implications. Meanwhile, there have been reports of humanlike biases in pretrained language models that have raised some valid concerns about how learners would be fed pre-formed judgments, whether in feedback or in

generation, if the language models were used in classrooms without proper critical scrutiny (Schramowski et al., 2022).

These pedagogically appealing features, such as generating fluent text, correcting errors in context, and responding in conversation, are rooted in the innovations in how LLM models are trained and fine-tuned. The innovations in the training and fine-tuning of LLM models are what give these models the features that make them pedagogically appealing: generating fluent text, correcting errors in context, and responding in conversation. Methods to remove duplicate training data have been demonstrated to have a measurable impact on the quality of downstream models and have the potential to mitigate memorization artifacts, an issue that is pertinent to the trustworthiness of AI-generated instructional content (Si et al., 2019; Lee et al., 2022). Likewise, chain-of-thought prompting has significantly boosted the performance of large models on multi-step reasoning tasks, a skill that is immediately applicable to producing pedagogically rich explanations as opposed to simple responses (Wei et al., 2022). More comprehensive studies of AI in education place these technologies in a wider context of AI-supported teaching and learning, and warn that a technology's ability to work well does not necessarily mean that it will be pedagogically effective (Holmes & Tuomi, 2022).

In addition to the quality of the raw output, the reliability of LLM outputs is significant for use in classrooms. Confidence calibration in language models, where AI systems dynamically decide which aspects of output to allocate more resources to and which to neglect, suggests methods that could be used by teachers to identify confidently generated feedback versus output that demands more attention (Schuster et al., 2022). However, the human aspect of English teaching still plays a significant role in this discussion, as research on teachers' professional identity and autonomy indicates that the adoption of technology is shaped by the teachers' sense of instructional ownership, which can either facilitate or hinder the acceptance or rejection of AI tools, and selective use in class (Derakhshan et al., 2020). Additionally, ongoing research in auto formalization, the transformation of informal language into structured and precise representations, indicates that LLMs could be used for more stringent automated language assessment of learners that goes beyond checking grammatical structure (Wu et al., 2022). Lastly, research on vocabulary learning from input (in-class and extra-class exposure) can be used to establish a baseline to meaningfully assess the incremental effect of LLM-supported input, including AI-generated reading materials or conversational practice (Peters et al., 2019).

By proposing a concept framework incorporating the application of LLMs into ELT, this study is part of the research that investigates the use of LLMs in ELT, with the focus on teacher supervision as the core design principle of using LLMs in ELT. It offers empirical evidence from a ten-week quasi-experimental pilot that examined LLM-supported teaching practices against traditional approaches to lesson planning and feedback that do not utilize AI. The study also highlights practical implications such as the potential for overreliance, the consistency of outputs, and the impact of AI tools on educators' professional identity. Moreover, it provides organized evaluation metrics and teacher–learner outcome measures and offers a basis for future comparative studies regarding the effectiveness and responsible use of English language learning assisted by LLMs.

The rest of this paper is organized as follows. The scholarship literature in the areas of LLM capabilities and applications in ELT is reviewed in Section II. The conceptual framework of the proposed teaching support system based on LLM is introduced in Section III. The research design used in a pilot evaluation is described in Section IV. Classroom integration (CI) workflow in the study is described in Section V. The evaluation results of the pilot are presented in Section VI. The paper ends in a summary section VII, in which insights and directions for future research are summarized.

Large Language Model Capabilities and Relevance to English Language Teaching

A body of research has emerged showing that the technical capabilities of the large language models are found in functions that have direct applicability in the classroom. In general, transformer-based language models can be adapted to uncover the exact meaning of a writer's ideas, which has been demonstrated for the task of clinical documentation, and may be transferred to automated feedback about grammar and usage from learner writing. To complement this technical base, research on intelligent, AI-powered English teaching platforms shows that adaptive platforms integrating content suggestion and automatic assessment into the classroom can be effectively used in teaching and learning activities when added on top of teacher-led activities, instead of replacing teachers (Sun et al., 2021; Elazar et al., 2021).

A major limitation on using pretrained language models for classroom purposes is that models may not consistently answer logically equivalent questions, with direct relevance to the use of AI-generated explanations of grammar rules, which may be confusing to students if not consistent in phrasing question. A major limitation on the use of pretrained language models for classroom purposes is that models do not consistently answer logically equivalent questions, which has direct implications for teachers who use AI-generated explanations of grammar rules: if the same models can be inconsistent in phrasing, then learners may be confused by the model's answer, which is not the same as the teachers. Bangladesh's context-specific studies of English language teaching reveal that the goals of English language proficiency as set out in policy are often more ambitious than those reflected in the resources available in the classroom and the education of the teachers, a dynamic that is likely to repeat, if not intensify, when new technologies such as LLM are brought into similarly resource-poor educational contexts without the appropriate support systems (Rahman et al., 2019).

In addition to classroom applications, the study of virtual exchange as a teaching approach for language shows that technology-based interaction between learners from different institutions and countries can enable language learning to be an authentic communicative practice, indicating a complementary route by which learners' conversational practice with LLM could be integrated with authentic peer interaction (Wang, 2019). Research findings from translanguaging in the Chinese foreign language classroom also reveal that the FL teacher and learner tend to switch among various linguistic resources in foreign language teaching, which may pose a challenge for LLM systems that are mainly designed for monolingual target language output if not explicitly designed to support such practices in foreign language learning (Ofer et al., 2021).

Advances in technology outside the field of language teaching continue to shape the larger movement of LLM competence for teaching. The ability to adapt transformer-based architectures to different domains of symbols, as shown in the field of protein sequence analysis, highlights remarkable versatility and suggests that well suited for specific education-related tasks like automated writing assessment (Dooly & Vinagre, 2022; Naseem et al., 2021). In-depth studies of word representations from traditional embedding methods to the use of contextualized representations in today's LLM systems offer valuable technical insights for educators interested in the rationale behind the improvements of contemporary models over past rule-based grammar checkers (Zein et al., 2020).

Studies analyzing the situation of English education and research in Indonesia show how the latest technologies enter into the national research agenda after being successful in other countries, thus demonstrating that language teaching practices formed with the help of LLMs can quickly spread to similar educational environments (Haleem et al., 2022). Moreover, the studies dedicated to all ChatGPT-related

issues explore its potential as a futuristic tool that might provide personalized educational assistance along with some critique related to the inaccuracy of information and ethical issues (Schmidt & Strasser, 2022).

The results of the current study suggest that different LLMs can be effectively used to facilitate the process of English language teaching when used in a framework involving a teacher controlling the process rather than producing the content independently. Thus, the decreased lesson preparation time and increased grammatical accuracy of students prove that the assistance of LLMs can improve personalized feedback while staying practical. Nevertheless, it is worth mentioning that teachers face certain challenges connected to the way perceive automation as evaluations indicate the importance of professional identity and responsible use of AI technologies. Moreover, findings show that teachers tend to analyze the content produced by LLMs and adjust it accordingly to specific purposes.

A Framework for Integrating Large Language Models into English Teaching Practice

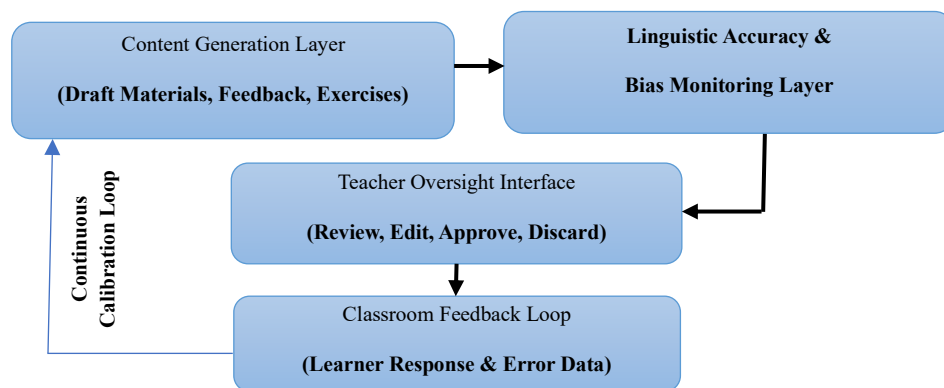


Figure 1: Architecture of the Proposed LLM-Integrated Teaching Support Framework

The framework for the design of the proposed system, depicted in figure 1, essentially consists of four layers. This is anchored by the teacher-centered oversight loop. The content generation layer provides draft instructional materials, which include reading passages and practice exercises, as well as personalized feedback comments according to parameters specified by the teacher, such as the learner's proficiency level and target linguistic features. Outputs from this layer are directed to the linguistic accuracy and bias monitoring layer, where the content is checked against factual inaccuracies, inconsistencies, and bias before it is presented to either the instructor or the learner.

The output of the screening process is sent to the teacher oversight interface, which serves as the focal structural component of the framework. Teachers are able therefore to consider and amend AI-generated materials before approving or rejecting use in classes: the final decision of what instructional materials to be used belongs to the teacher rather than the model. Once these materials are accepted and put into the feedback loop, the learners are instructed and receive feedback input as well as own responses on the provided input so that the model is able to further improve its accuracy and make the generated content more suitable for the specific target audience. In this way, the closed-loop system deliberately ensures that the teacher actively participates in the process, either approving or rejecting the AI-generated materials.

Research Design for Examining LLM-Assisted English Teaching

This research used a mixed-method, quasi-experimental design, which incorporated both framework application and school-based investigation, thus enabling not only quantitative assessment of instructional results but also qualitative knowledge of teachers' experience of LLM adaptation.

Participants included thirty-six teachers of English from secondary school and adult-learning programs and learners who were enrolled in classes. Teachers were assigned to one of two groups: the experimental group ($n = 18$) that used the developed LLM-based teaching support system and the control group ($n = 18$) that continued using traditional methods of lesson planning and teaching for ten weeks. All teachers participating in the study had a minimum of three years in teaching and had volunteered to participate in the research after getting information about the study.

Data were collected through multiple instruments with a triangulation purpose. Teacher surveys were administered before and after the intervention to reveal how teachers estimated the time spent on lesson preparation, the frequency of feedback, and confidence in addressing the needs of various groups of learners. Writing samples were collected to assess grammatical accuracy and quality of revisions. The lesson observations showed how the materials generated by LLMs were used, changed, or ignored by experimental teachers. Twelve teachers were interviewed to gather opinions regarding the framework.

Quantitative data obtained from the surveys and writing samples were analyzed using descriptive statistics. Data collected during interviews and observations were evaluated using thematic coding. The combination of quantitative and qualitative approaches gave a more comprehensive view of instructional results obtained and teachers' impressions about the use of LLM. Ethics approval for the study was obtained before the data collection. Informed consent for the research was gained from the participants.

Classroom Integration and Adaptive Instructional Workflow

The framework in question was developed into an online platform for teacher use, where would be able to request lesson drafts and comments, modify and finalize the results of AI processing, and send approved materials via the educational institution's LMS. Figure 2 illustrates the conceptual structure of the system that provides data from the educational setting for progressive AI-generated learning.

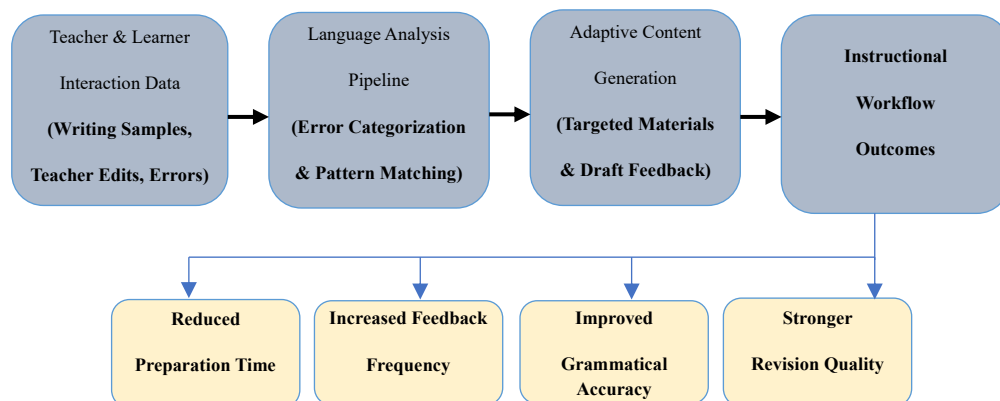


Figure 2: LLM-Assisted Feedback and Lesson Planning Workflow

The data on interactions between learners and instructors involving submitted written works, instances of educator alteration of responses generated by the AI as well as error patterns is shown in figure 2. The system looks into all kinds of grammatical and lexical errors and compares them against any previously

noted errors for either the same student or the same class. The information identified is then used to create adaptive content since it is based on those unique errors made by the learners, allowing the system to generate feedback and suggest adaptive practical tasks, then verified by the instructor before being sent.

Changes made by the educators to the AI-generated remarks at this stage, be it corrections and adjustments made by the teachers, are also logged into the system contributing to making changes in the future for that educator's students. To avoid total dependence on the AI during the whole work, many measures were incorporated, including a mandatory review step before any feedback is passed down to the learners, as well as the teachers being periodically prompted to compare AI-generated feedback with own.

Teacher and Learner Outcomes in LLM-Assisted English Instruction

The assessment of instructional outcomes was done by contrasting pre-and post-intervention survey responses and writing scores of the learners from both educational groups, while qualitative data were supplied by the participating teachers. The average calculation of the data received from the teachers working with experimental groups has shown that there was a decrease of thirty-five percent in the time spent on lesson preparations and an eight percent decrease in lesson-preparation time in the comparison group, while the level of grammatical accuracy of students' assignments was higher by eighteen percentage points in the experimental educational groups rather than nine percentage points in the control group.

Table 1: Teachers' Perceptions of the LLM-Assisted English Teaching Framework

Statement	Agree (%)	Neutral (%)	Disagree (%)
The framework reduced weekly lesson-preparation workload	81	12	7
They were able to give more frequent individualized feedback to learners	84	10	6
Generated materials were appropriate to learners' proficiency level	76	16	8
Using the framework changed how confident they feel evaluating learner work independently	52	29	19
They would recommend this framework to other English teachers	79	14	7

Table 1 presents the outcomes of the survey carried out among the teachers regarding the LLM-based English teaching framework. The majority of respondents agreed that the framework decreased the amount of work for preparation, encouraged more frequent personalized feedback, and offered resources that corresponded with students' level of knowledge. There was a lower degree of agreement regarding the issue of whether or not the framework affected teachers' confidence in the evaluation of students' work, a point that will be analyzed in detail in chapter VII.

Observation of the lessons showed that the teachers from the experimental group often used machine-generated materials as a draft text that edited significantly instead of passing it to students unaltered. The nature of changes was more about tone and culture than about grammatical correctness. An analysis of the interview data revealed that teachers appreciated time saved on routine feedback; however, not sure about faith in the machine-generated explanations of grammar methods that need to be double-checked independently. Some teachers indicated that the obligatory review of the materials was sometimes experienced as additional work but still reinforced professional autonomy in the contents of the lesson.

Conclusion and Future Directions

This study investigated the ways that the large language models can be strategically integrated into English instruction while retaining teacher supervision; and the outcomes from the research provide promising, but appropriately restrained, conclusions about the approach discussed. According to pilot evaluation data, the framework proposed reduced preparation time of teachers and increased the frequency of feedback provided by teachers compared to the opponents, while learners achieved better results compared to peers regarding accuracy and quality of revisions made. While meeting the criteria outlined by the literature on the impact of automation on professional judgment, the four-level architecture of the framework, which includes content generation and monitoring linguistic biases, instructor supervision, and feedback collection in the classroom contributed to the effectiveness of the method.

However, it should be noted that the study revealed a few questions to be considered in the further implementation. The questions include bias monitoring, realistic expectations regarding the reliability of the models used, and how the use of AI influences the identity of teachers. Future studies must be aimed to study the topic further through the extended longitudinal studies, bigger social context, and comparative studies of architects and supervision approaches used in various large language models. Moreover, it is important to consider learners and create means aiming to provide teachers with the chance to assess AI instruction materials critically. Thus, the study emphasizes the fact that large language models can effectively be used in instruction in case are integrated around teacher supervision rather than replacement.

References

- Derakhshan, A., Coombe, C., Arabmofrad, A., & Taghizadeh, M. (2020). Investigating the effects of English language teachers' professional identity and autonomy in their success. *Issues in Language Teaching*, 9(1), 1-28. <https://doi.org/10.22054/ilt.2020.52263.496>
- Dooly, M., & Vinagre, M. (2022). Research into practice: Virtual exchange in language teaching and learning. *Language Teaching*, 55(3), 392-406. <https://doi.org/10.1017/S0261444821000069>
- Elazar, Y., Kassner, N., Ravfogel, S., Ravichander, A., Hovy, E., Schütze, H., & Goldberg, Y. (2021). Measuring and improving consistency in pretrained language models. *Transactions of the Association for Computational Linguistics*, 9, 1012-1031. https://doi.org/10.1162/tacl_a_00410
- Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil transactions on benchmarks, standards and evaluations*, 2(4), 100089. <https://doi.org/10.1016/j.tbench.2023.100089>
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European journal of education*, 57(4), 542-570. <https://doi.org/10.1111/ejed.12533>
- Lee, K., Ippolito, D., Nystrom, A., Zhang, C., Eck, D., Callison-Burch, C., & Carlini, N. (2022, May). Deduplicating training data makes language models better. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 8424-8445). <https://doi.org/10.18653/v1/2022.acl-long.577>
- Naseem, U., Razzak, I., Khan, S. K., & Prasad, M. (2021). A comprehensive survey on word representation models: From classical to state-of-the-art word representation language models. *Transactions on Asian and Low-Resource Language Information Processing*, 20(5), 1-35. <https://doi.org/10.1145/3434237>

- Ofer, D., Brandes, N., & Linial, M. (2021). The language of proteins: NLP, machine learning & protein sequences. *Computational and Structural Biotechnology Journal*, 19, 1750-1758. <https://doi.org/10.1016/j.csbj.2021.03.022>
- Peters, E., Noreillie, A. S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language learning*, 69(3), 747-782. <https://doi.org/10.1111/lang.12351>
- Rahman, M. M., Islam, M. S., Karim, A., Chowdhury, T. A., Rahman, M. M., Seraj, P. M. I., & Singh, M. K. M. (2019). English language teaching in Bangladesh today: Issues, outcomes and implications. *Language Testing in Asia*, 9(1), 9. <https://doi.org/10.1186/s40468-019-0085-8>
- Schmidt, T., & Strasser, T. (2022). Artificial intelligence in foreign language learning and teaching: a CALL for intelligent practice. *Anglistik: International Journal of English Studies*, 33(1), 165-184. <https://doi.org/10.33675/angl/2022/1/14>
- Schramowski, P., Turan, C., Andersen, N., Rothkopf, C. A., & Kersting, K. (2022). Large pre-trained language models contain human-like biases of what is right and wrong to do. *Nature Machine Intelligence*, 4(3), 258-268. <https://doi.org/10.1038/s42256-022-00458-8>
- Schuster, T., Fisch, A., Gupta, J., Dehghani, M., Bahri, D., Tran, V., ... & Metzler, D. (2022). Confident adaptive language modeling. *Advances in Neural Information Processing Systems*, 35, 17456-17472. <https://doi.org/10.52202/068431-1269>
- Si, Y., Wang, J., Xu, H., & Roberts, K. (2019). Enhancing clinical concept extraction with contextual embeddings. *Journal of the American Medical Informatics Association*, 26(11), 1297-1304. <https://doi.org/10.1093/jamia/ocz096>
- Sun, Z., Anbarasan, M., & Praveen Kumar, D. J. C. I. (2021). Design of online intelligent English teaching platform based on artificial intelligence techniques. *Computational Intelligence*, 37(3), 1166-1180. <https://doi.org/10.1111/coin.12351>
- Topsakal, O., & Topsakal, E. (2022). Framework for a foreign language teaching software for children utilizing AR, voicebots and ChatGPT (large language models). *The Journal of Cognitive Systems*, 7(2), 33-38. <https://doi.org/10.52876/jcs.1227392>
- Wang, D. (2019). Translanguaging in Chinese foreign language classrooms: Students and teachers' attitudes and practices. *International Journal of Bilingual Education and Bilingualism*, 22(2), 138-149. <https://doi.org/10.1080/13670050.2016.1231773>
- 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. <https://doi.org/10.52202/068431-1800>
- Wu, Y., Jiang, A. Q., Li, W., Rabe, M., Staats, C., Jamnik, M., & Szegedy, C. (2022). Autoformalization with large language models. *Advances in neural information processing systems*, 35, 32353-32368. <https://doi.org/10.52202/068431-2344>
- Zein, S., Sukyadi, D., Hamied, F. A., & Lengkanawati, N. S. (2020). English language education in Indonesia: A review of research (2011–2019). *Language Teaching*, 53(4), 491-523. <https://doi.org/10.1017/S0261444820000208>