Global Spectrum of Research and Humanities
ISSN: 3007-9136
Vol. 3 No. 1 (2026): Spring

Transformer Attention for Personalized L2 Feedback in CALL

[1] **Zarifa Sadigzade,** [2] **Hasan Alisoy**
https://doi.org/10.69760/gsrh.0260301006

Abstract; The rapid diffusion of transformer-based natural language processing has renewed long-standing ambitions in computer-assisted language learning (CALL): scalable, timely, and instructionally meaningful second language (L2) feedback that is sensitive to individual learners' needs rather than "one-size-fits-all" correction. Contemporary studies on chatbots and large language models (LLMs) suggest measurable promise for L2 vocabulary development, writing support, and learner engagement, while also revealing variability in feedback quality, inconsistency across prompts, and unresolved questions about pedagogical validity and accountability. This article synthesizes research from CALL, feedback theory, and transformer interpretability to propose an attention-informed framework for personalized L2 feedback. The framework treats transformer attention as a computational mechanism for aligning learner language with contextual cues and as a design resource for building inspectable feedback workflows, while explicitly recognizing that attention weights are not automatically faithful explanations of model behavior. Methodologically, we conduct a targeted, citation-grounded synthesis seeded by recent CALL chatbot/LLM research and complemented by foundational work on formative feedback and corrective feedback effectiveness, learner modeling, and transformer visual analytics. Results are presented as design-relevant findings: what current evidence implies for personalization targets (what to correct, how, when), how attention signals may support fine-grained diagnosis and recommendation, and which safeguards are required for responsible deployment in educational settings. Implications are discussed for research design (comparative trials against teacher feedback, delayed posttests, process data), system architecture (learner models + transformer feedback generators), and governance (privacy, transparency, and human oversight).
Keywords; *transformer attention, personalization, written corrective feedback, chatbots, large language models, learner modeling, explainability, CALL*

Introduction

Computer-assisted language learning has repeatedly returned to a central instructional challenge: learners benefit from feedback that is timely, specific, and appropriately scaffolded, yet instructor-generated feedback is time-intensive and difficult to sustain at scale—especially for extended writing and iterative revisions (Hattie & Timperley, 2007; Shute, 2008; Ferris, 2010). Large language models and AI writing tools, typically built on transformer architectures, have shifted the feasibility frontier for automated feedback because they can generate context-sensitive language, propose corrections, and simulate dialogic interaction (Godwin-Jones, 2022; Kasneci et al., 2023;

[1] Zarifa Sadiqzade, lecturer, Nakhchivan State University, zarifasadig@gmail.com, ORCID: https://orcid.org/0009-0007-1179-1214

[2] Alisoy, H. Lecturer in English, Nakhchivan State University, Azerbaijan. Email: alisoyhasan@ndu.edu.az. ORCID: https://orcid.org/0009-0007-0247-476X

Mizumoto & Eguchi, 2023). However, evidence from L2 contexts indicates that automated feedback can differ from teacher feedback in form and consistency, and may sometimes be redundant, unstable across repeated prompts, or misaligned with instructional priorities (Lin & Crosthwaite, 2024).

From an L2 acquisition perspective, the concept of "attention" has pedagogical meaning: noticing and awareness are widely treated as mechanisms that mediate how input becomes intake, and how learners respond to feedback and form-focused instruction (Schmidt, 1990; Ellis, 2009). In parallel, in modern NLP the term "attention" also denotes a calculable mechanism—self-attention—that allows transformer models to represent relationships among tokens in a sequence, enabling strong performance across language tasks (Vaswani et al., 2017). The opportunity for CALL is not merely to apply transformers, but to connect computational attention to pedagogical objectives: selecting feedback targets, calibrating feedback explicitness, and sequencing follow-up practice with sensitivity to the learner's trajectory.

Yet a critical caveat shapes this agenda: attention weights are often visually compelling and intuitively "explainable," but empirical work shows that attention is not automatically a faithful explanation of model decisions; different attention patterns can sometimes yield similar predictions, and high attention weights do not guarantee causal influence (Jain & Wallace, 2019; Serrano & Smith, 2019). Accordingly, using attention in CALL requires a principled distinction between attention as (a) a modeling mechanism and (b) an interpretation or explanation artifact for teachers and learners (Wiegreffe & Pinter, 2019).

The present article addresses the following research problem: How can transformer attention be operationalized to support personalized L2 feedback that is pedagogically credible, empirically evaluable, and transparent enough for classroom adoption? To answer, we synthesize: (a) CALL evidence on chatbots/LLMs for language learning, (b) feedback and corrective feedback research relevant to design constraints, and (c) transformer attention scholarship (architecture, visualization, and interpretability).

Methods

This article follows a targeted research synthesis + design framework methodology appropriate for a fast-evolving sociotechnical domain where educational evaluation and computational capabilities co-develop. The method has three stages.

First, we assembled a seed corpus from recent CALL and educational technology research on chatbots, LLMs, and automated feedback provided by the prompt (e.g., Godwin-Jones, 2022; Huang et al., 2022; Jeon, 2023; Kim et al., 2022; Lin & Crosthwaite, 2024; Liu et al., 2024; Mizumoto & Eguchi, 2023; Yuan, 2024; Zhai & Wibowo, 2022; Zhang & Huang, 2024) and from empirical conversational learning systems (Ruan et al., 2019, 2021). We then expanded the corpus through iterative searches for foundational and bridge literature on (a) formative feedback and corrective feedback effectiveness, (b) transformer architecture and attention interpretation, and (c) learner modeling and personalization mechanisms.

Second, we conducted conceptual coding aimed at design translation rather than exhaustive meta-analysis. Each source was coded for: (a) feedback function (error identification, explanation, strategy instruction, affective support), (b) personalization target (proficiency, error profiles, motivation, SRL), (c) interaction format (dialogue, annotation, recommendation), and (d) evaluation evidence (learning gains, delayed retention, process measures). This coding aligns with established distinctions in feedback scholarship—e.g., feedback can support learning but its effects depend on task, timing, and information level (Kluger & DeNisi, 1996; Hattie & Timperley, 2007; Shute, 2008)—and with L2 writing corrective feedback typologies that differentiate direct/indirect/metalinguistic strategies and required learner responses (Ellis, 2009; Ferris, 2010; Kang & Han, 2015).

Third, we used the coded evidence to propose a Transformer Attention for Personalized Feedback (TAPF) framework. TAPF is presented as an implementable architecture and an evaluation agenda, grounded in current transformer practice. The architecture assumes (a) a transformer-based language model for error detection and feedback generation (Vaswani et al., 2017; Omelianchuk et al., 2020), (b) a learner model updated from interaction traces and performance signals (Corbett & Anderson, 1995; Piech et al., 2015), and (c) an adaptation layer that selects feedback type and follow-up practice using sequential personalization methods related to transformer-based recommendation (Sun et al., 2019; Wu et al., 2020). The framework also incorporates interpretability constraints by design: attention can be visualized and audited, but should be triangulated with other evidence before being treated as an explanation (Jain & Wallace, 2019; Serrano & Smith, 2019; Yeh et al., 2023).

Results

The synthesis yielded four design-relevant findings that jointly motivate transformer-attention–based personalization.

A first finding is that CALL chatbots and LLM-based tools show consistent potential for language learning, but outcomes and mechanisms vary by design and context. Systematic reviews of chatbot-supported language learning identify technological and pedagogical affordances (e.g., accessibility, interactional practice), while also reporting challenges such as limited social presence, constrained pedagogy, and uneven evaluation quality (Huang et al., 2022). Empirical studies show that chatbots can support vocabulary development and retention: in a controlled study, chatbot-assisted groups demonstrated improved receptive and productive vocabulary outcomes relative to control conditions (Zhang & Huang, 2024). Chatbots have also been shown to support dialogue-based learning and engagement: QuizBot improved recognition/recall relative to flashcards and increased voluntary time-on-task, suggesting motivational and interactional advantages of conversational practice formats (Ruan et al., 2019). In L2 contexts, however, design must account for learner affect and sociocultural appropriateness; cross-cultural, humor, and empathy dimensions are specifically highlighted as underdeveloped yet consequential (Zhai & Wibowo, 2022).

A second finding is that feedback effectiveness depends on form, focus, and learner engagement, implying that "more correction" is not equivalent to "better learning." In general educational research, feedback has strong potential influence on learning, but can also be ineffective or negative depending on how it is framed and what it targets (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). In formative feedback research, effectiveness is associated with qualities such as specificity, timeliness, and credibility, and with feedback that supports improvement rather than merely evaluation (Shute, 2008). In L2 contexts, meta-analytic evidence indicates that corrective feedback can improve linguistic accuracy, with outcomes moderated by methodological and contextual factors (Li, 2010). For written corrective feedback specifically, meta-analysis supports positive effects on L2 written accuracy while emphasizing mediators and moderators (Kang & Han, 2015). These findings constrain the design space for transformer-based feedback: personalization must include decisions about what to correct (e.g., targeted structures), how explicit to be (direct vs. metalinguistic), and how to induce uptake (revision requirements, practice).

A third finding is that LLM-mediated L2 feedback is empirically promising but demonstrably different from teacher feedback, which raises alignment and accountability issues. Work comparing teacher-provided written corrective feedback with GPT-assisted feedback reports that teacher feedback and GPT feedback differ in typical forms and emphases; GPT-assisted feedback may provide metalinguistic commentary or reformulation and can vary its approach even under identical prompts (Lin & Crosthwaite, 2024). In parallel, experimental integration of LLM support into EFL writing instruction has been associated with improvements in writing performance and learning-related variables such as self-regulated learning strategies and motivation, indicating that LLMs can be embedded in more comprehensive pedagogical designs rather than used only as "instant correction" (Liu et al., 2024). In assessment-adjacent domains, LLMs have been investigated for automated essay scoring with evidence that such systems can achieve useful levels of accuracy and reliability while still raising questions about validity and the role of linguistic feature controls (Mizumoto & Eguchi, 2023). Taken together, the literature suggests that transformer-based feedback should be treated as instructional infrastructure requiring teacher mediation and curricular integration, not as a fully autonomous evaluator or instructor (Godwin-Jones, 2022; Kasneci et al., 2023).

A fourth finding is that transformer attention is simultaneously enabling and methodologically risky as a basis for personalization and explanation. The transformer architecture uses self-attention to model contextual dependencies and has become foundational for modern language models (Vaswani et al., 2017; Devlin et al., 2019). Attention can be inspected and visualized, and a growing body of visualization work aims to support global understanding of attention patterns across inputs (Yeh et al., 2023) as well as more localized exploration of multi-head attention (Vig, 2019). Nevertheless, interpretability research cautions against equating attention weights with faithful explanation of model decisions (Jain & Wallace, 2019; Serrano & Smith, 2019). Counter-arguments distinguish between "plausibility" and "faithfulness," suggesting that attention may be useful for certain explanatory goals under explicit tests and constraints (Wiegreffe & Pinter, 2019). For personalized L2 feedback, this produces a concrete design implication: attention can provide candidate alignment signals (e.g., which tokens are associated with suggested corrections), but pedagogical transparency should rely on triangulation—error-type labeling, rule-based checks, or counterfactual tests—rather than attention heatmaps alone.

Discussion

The evidence supports a clear conclusion: transformer-based systems can plausibly deliver personalized L2 feedback if personalization is treated as a multi-layer design problem spanning learner modeling, feedback typology, sequencing, and governance. This section articulates TAPF as an actionable agenda for CALL research and practice.

At the architectural level, TAPF links three components. The first is an L2 feedback engine that performs grammatical error correction, fluency editing, and explanation generation. Transformer encoders can support efficient correction strategies such as tagging-based edits, which are computationally suitable for interactive environments that require low latency (Omelianchuk et al., 2020). For fluency-sensitive feedback, datasets and benchmarks from grammatical error correction research emphasize that "native-like" revision may involve more than local grammar fixes, reinforcing the need to distinguish accuracy correction from stylistic reformulation in pedagogical interfaces (Napoles et al., 2017).

The second component is a learner model that updates over time from learner performance, preferences, and interaction traces. Knowledge tracing (Corbett & Anderson, 1995) and neural variants (Piech et al., 2015) illustrate how systems can estimate changing mastery states during learning. In CALL, such learner models can represent recurring error patterns (e.g., article use, tense marking), response to feedback types (direct vs. indirect), and time-varying engagement signals. The third component is an adaptation and recommendation layer that decides which feedback to present and which tasks to assign next. Transformer-based sequential recommendation approaches show how attention can model evolving preferences and histories—an analogy directly transferable to sequences of learner interactions and error events (Sun et al., 2019; Wu et al., 2020).

At the pedagogical level, attention must be grounded in what L2 feedback is supposed to accomplish. Feedback theory emphasizes that feedback effectiveness is conditional: it can help learners close gaps relative to goals, but may also misfire if it shifts focus away from the task or triggers maladaptive self-evaluation (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). In L2 writing, corrective feedback is not monolithic; typologies distinguish direct correction, indirect marking, and metalinguistic explanation, and these choices interact with learner revision behavior and instructional purpose (Ellis, 2009; Ferris, 2010). TAPF therefore treats personalization as feedback routing: for a given learner at a given time, the system chooses among feedback modes (e.g., minimal hint vs. explicit correction), determines whether revision is required, and selects follow-up practice that targets the learner's current zone of development (Vygotsky, 1978). This is conceptually compatible with dynamic assessment approaches in which graduated prompts serve both intervention and diagnosis; evidence from chatbot-assisted dynamic assessment shows that automated scaffolding can yield higher vocabulary gains while producing diagnostic interaction records (Jeon, 2023).

At the explainability level, TAPF recommends an "attention-informed but not attention-only" stance. Visual analytics work suggests that attention can be explored at scale and may help developers and researchers detect global attention patterns (Yeh et al., 2023), while tools like BertViz illustrate how attention views can support model analysis (Vig, 2019). Yet interpretability research indicates that attention weights can be misleading as causal explanations (Jain & Wallace, 2019; Serrano & Smith, 2019). For CALL, this means attention visualizations should be used primarily for (a) system diagnostics and teacher-facing auditing and (b) learner-facing plausibility cues paired with pedagogical labels (e.g., "article choice," "verb form," "collocation") and actionable revision prompts. This approach aligns with classroom realities documented in comparisons of teacher and GPT feedback: teachers' feedback blends local and global concerns with a degree of inaccuracy, while GPT feedback may be grammatically correct yet pedagogically redundant or inconsistent, implying that transparency and teacher mediation are central for practical adoption (Lin & Crosthwaite, 2024).

At the evaluation level, the literature supports moving beyond satisfaction surveys toward mixed evidence: learning gains, delayed retention, uptake, and interaction traces. Recent L2 chatbot studies provide examples of controlled designs with delayed posttests (Jeon, 2023; Zhang & Huang, 2024), while writing-support studies point to the relevance of motivational and self-regulation outcomes when LLMs are integrated into instruction (Liu et al., 2024). TAPF therefore recommends evaluation packages that include: (a) mastery outcomes (accuracy, fluency, complexity) with delayed measures, (b) feedback uptake (revision quality, error recurrence), and (c) process data (turn-by-turn interaction, time-on-task), consistent with prior evidence that conversational agents can change engagement patterns, not just test scores (Ruan et al., 2019).

Finally, at the governance level, responsible personalization requires explicit safeguards. Major syntheses and guidance documents emphasize risk landscapes for foundation models used in education and the need for human-centered policy and practice, including attention to privacy, bias, and misuse (UNESCO, 2023; Zawacki-Richter et al., 2019; Bommasani et al., 2021; Bender et al., 2021; Weidinger et al., 2021). In CALL settings, these concerns become concrete: learner texts can contain personal data, feedback can shape learner identity and confidence, and conversational interfaces can create over-trust. Accordingly, TAPF treats teacher oversight, data minimization, and transparency about system limits as design requirements rather than optional add-ons. This helps reconcile the promise of scalable feedback with the ethical obligation to protect learners and maintain instructional integrity (Godwin-Jones, 2022; Kasneci et al., 2023).

References

AbuSahyon, A. S. A. E., Alzyoud, A., Alshorman, O., & Al-Absi, B. (2023). AI-driven technology and chatbots as tools for enhancing English language learning in the context of second language acquisition: A review study. International Journal of Membrane Science and Technology, 10(1), 1209–1223.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623).

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., … Liang, P. (2021). On the opportunities and risks of foundation models. arXiv.

Brusilovsky, P. (2001). Adaptive hypermedia. User Modeling and User-Adapted Interaction, 11, 87–110.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction, 4, 253–278.

Debets, T., Claes, J., & Vanderhoven, E. (2025). Chatbots in education: A systematic review of objectives, characteristics, and evaluation methods. Computers & Education, 204, 105055.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 4171–4186).

Ellis, R. (2009). A typology of written corrective feedback types. ELT Journal, 63(2), 97–107.

Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. Studies in Second Language Acquisition, 32(2), 181–201.

Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. Language Learning & Technology, 26(2), 5–24.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112.

Heift, T., & Schulze, M. (2007). Errors and intelligence in computer-assisted language learning: Parsers and pedagogues. Routledge.

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. Journal of Computer Assisted Learning, 38(1), 237–257.

Huang, Y., Geertzen, J., Baker, R., Korhonen, A., & Alexopoulou, T. (2017). The EF-Cambridge Open Language Database (EFCamDat): Information for users (Second release).

Jain, S., & Wallace, B. C. (2019). Attention is not explanation. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 3543–3556).

Jeon, J. (2023). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. Computer Assisted Language Learning, 36(7), 1338–1364.

Kang, E. Y., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. The Modern Language Journal, 99(1), 1–18.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Kuhn, J. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.

Kim, H., Yang, H., Shin, D., & Lee, J. H. (2022). Design principles and architecture of a second language learning chatbot. Language Learning & Technology, 26(1), 1–18.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin, 119(2), 254–284.

Kohnke, L. (2023). L2 learners' perceptions of a chatbot as a potential independent language learning tool. International Journal of Mobile Learning and Organisation, 17(1–2), 214–226.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. Language Learning, 60(2), 309–365.

Lin, S., & Crosthwaite, P. (2024). The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback. System, 127, 103529.

Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. AI Open, 3, 111–132.

Liu, Z. M., Hwang, G. J., Chen, C. Q., Chen, X. D., & Ye, X. D. (2024). Integrating large language models into EFL writing instruction: Effects on performance, self-regulated learning strategies, and motivation. Computer Assisted Language Learning, 1–25.

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. Research Methods in Applied Linguistics, 2(2), 100050.

Napoles, C., Sakaguchi, K., & Tetreault, J. (2017). JFLEG: A fluency corpus and benchmark for grammatical error correction. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (pp. 229–234).

Omelianchuk, K., Atrasevych, V., Chernodub, A., & Skurzhanskyi, O. (2020). GECToR—Grammatical error correction: Tag, not rewrite. In Proceedings of the 15th Workshop on Innovative Use of NLP for Building Educational Applications (pp. 163–170).

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In Advances in Neural Information Processing Systems.

Prasad, P., & Sane, A. (2024). A self-regulated learning framework using generative AI and its application in CS educational intervention design. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education (Vol. 1, pp. 1070–1076).

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A primer in BERTology: What we know about how BERT works. Transactions of the Association for Computational Linguistics, 8, 842–866.

Ruan, S., Jiang, L., Xu, J., Tham, B. J. K., Qiu, Z., Zhu, Y., & Landay, J. A. (2019). QuizBot: A dialogue-based adaptive learning system for factual knowledge. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–13).

Ruan, S., Jiang, L., Xu, Q., Liu, Z., Davis, G. M., Brunskill, E., & Landay, J. A. (2021). EnglishBot: An AI-powered conversational system for second language learning. In Proceedings of the 26th International Conference on Intelligent User Interfaces (pp. 434–444).

Schmidt, R. W. (1990). The role of consciousness in second language learning. Applied Linguistics, 11(2), 129–158.

Serrano, S., & Smith, N. A. (2019). Is attention interpretable? In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 2931–2951).

Shute, V. J. (2008). Focus on formative feedback. Review of Educational Research, 78(1), 153–189.

Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., & Jiang, P. (2019). BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (pp. 1441–1450).

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems.

Vig, J. (2019). BertViz: A tool for visualizing multi-head self-attention in transformer models. In ICLR Workshop on Debugging Machine Learning Models.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., … Gabriel, I. (2021). Ethical and social risks of harm from language models. arXiv.

Wiegreffe, S., & Pinter, Y. (2019). Attention is not not explanation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (pp. 11–20).

Wu, H., Shao, Y., Ozfatura, E., Mikolajczyk, K., & Gündüz, D. (2024). Transformer-aided wireless image transmission with channel feedback. IEEE Transactions on Wireless Communications, 23(9), 11904–11919.

Wu, L., Leung, H. F., & Luong, H. C. (2017). Design and analysis of CMOS LNAs with transformer feedback for wideband input matching and noise cancellation. IEEE Transactions on Circuits and Systems I: Regular Papers, 64(6), 1626–1635.

Wu, L., Li, S., Hsieh, C.-J., & Sharpnack, J. (2020). SSE-PT: Sequential recommendation via personalized transformer. In Proceedings of the 14th ACM Conference on Recommender Systems (pp. 328–337).

Yeh, C., Chen, Y., Wu, A., Chen, C., Viégas, F., & Wattenberg, M. (2023). AttentionViz: A global view of transformer attention. arXiv.

Yin, Q., & Satar, M. (2020). English as a foreign language learner interactions with chatbots: Negotiation for meaning. International Online Journal of Education and Teaching, 7(2), 390–410.

Yuan, Y. (2024). An empirical study of the efficacy of AI chatbots for English as a foreign language learning in primary education. Interactive Learning Environments, 32(10), 6774–6789.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? International Journal of Educational Technology in Higher Education, 16, 39.

Zhai, C., & Wibowo, S. (2022). A systematic review on cross-culture, humor and empathy dimensions in conversational chatbots: The case of second language acquisition. Heliyon, 8(12), e12056.

Zhang, Z., & Huang, X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. Heliyon, 10(3), e25370.

Received: 20.01.2026
Revised: 02.01.2026
Accepted: 08.02.2026
Published: 12.02.2026