



The Use of Large Language Models in Foreign Language Education

Anna Shcherbakova¹, Andrey Shcherbakov^{2,3}

¹Independent Researcher, Marseille, France

²Department of Cognitive-Analytical and Neuro-Applied Technologies, Russian State Social University, Moscow, Russian

³Department of Mathematical Modeling, State University of Management, Haifa, Israel

Email: x509@ras.ru

How to cite this paper: Shcherbakova, A. and Shcherbakov, A. (2026) The Use of Large Language Models in Foreign Language Education. *Open Access Library Journal*, **13**: e15245.
<https://doi.org/10.4236/oalib.1115245>

Received: March 26, 2026

Accepted: April 13, 2026

Published: April 16, 2026

Copyright © 2026 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Large language models (LLMs), such as GPT-based systems, are transforming foreign language education by enabling personalized practice, instant feedback, and immersive interactions beyond traditional computer-assisted language learning (CALL) tools. The article outlines LLMs' advantages over rule-based systems, analyzes applications across listening, speaking, reading, and writing skills, and addresses risks like overreliance and bias. It emphasizes pedagogy-aligned integration to maximize benefits while mitigating ethical concerns. For listening and pronunciation, LLMs generate customized audio scripts via text-to-speech, improving comprehension but relying on quality TTS integration. In speaking, they simulate dialogues to build fluency and reduce anxiety, though they lack real-time pragmatics. Reading benefits from adaptive texts and glosses for vocabulary building, while writing leverages drafting and revision feedback, enhancing efficiency in large classes. Cognitive risks include shallow learning and assessment disruption; biases from training data affect cultural representation; privacy issues arise from data usage. Design principles advocate framing LLMs as supplements, embedding metacognitive scaffolds, clear guidelines, and teacher training. Learner attitudes are positive, especially for out-of-class practice, but effects vary by proficiency. LLMs hold promise for scalable language learning when critically structured, calling for longitudinal research on long-term competence.

Subject Areas

Linguistics

Keywords

Large Language Model, Artificial Intelligence, Computer-Assisted Language Learning (CALL)

1. Introduction

Large language models (LLMs) such as GPT-based systems have rapidly become salient tools in education, reshaping how learners interact with language and access instructional support [1] [2]. In foreign language education, these models offer new possibilities for personalized practice, immediate feedback, and immersive communication that go beyond earlier generations of computer-assisted language learning (CALL) systems [2] [3]. At the same time, their integration raises critical questions about pedagogical value, cognitive impact, and ethical risks, including over-reliance, bias, and threats to academic integrity [4] [5].

This article examines the use of LLMs in foreign language education with three objectives. First, it provides a conceptual overview of how LLMs differ from traditional CALL tools and rule-based or retrieval-driven systems. Second, it analyzes key application areas across the four core language skills (listening, speaking, reading, and writing), drawing on emerging empirical evidence. Third, it discusses the main risks and outlines design principles for responsible, pedagogy-aligned use of LLMs as language learning tools rather than as mere answer-generators [4] [6].

2. Conceptual Background: From CALL to LLM-Mediated Learning

Early CALL systems largely relied on predefined templates, limited pattern-matching, and static feedback mechanisms, which constrained the range and authenticity of learner-computer interaction [3]. By contrast, contemporary LLMs are trained on massive multilingual corpora and can generate contextually rich, open-ended responses, enabling more naturalistic dialogue, adaptive scaffolding, and on-the-fly task design [2] [7]. This shift from fixed scripts to generative interaction allows learners to explore diverse topics and communicative scenarios that would be difficult to pre-author manually.

However, the generative capacity of LLMs also introduces epistemic uncertainty: the same mechanism that enables flexible, creative output can produce hallucinations, subtle grammatical errors, or culturally inappropriate content [4] [5]. In the context of foreign language learning, this means that LLMs can simultaneously function as powerful tutors and unreliable informants, making pedagogical framing and critical digital literacy essential. Recent position and benchmark papers emphasize that LLMs should be conceptualized as fallible socio-technical tools whose pedagogical competence and limitations must be explicitly evaluated, rather than assumed [6] [8].

3. Applications Across Core Language Skills

Listening and Pronunciation

LLMs, when integrated into multimodal systems with text-to-speech capabilities, can generate varied audio materials tailored to learners' proficiency levels and interests [1] [3]. Systems built around LLM-generated scripts and synthetic voices

have been reported to improve learners' listening comprehension, especially when learners repeatedly engage with customized audio that reflects diverse accents and registers [3]. Such tools can also support pronunciation practice by providing immediate textual feedback on segmental accuracy and prosodic features based on the learner's recorded speech.

Nevertheless, high-quality pronunciation feedback still depends on the integration of robust speech recognition and phonetic analysis modules, which may lag behind the sophistication of text-based LLM components [3]. In addition, synthetic voices risk normalizing a narrow set of accent norms if not carefully diversified, potentially reinforcing existing hierarchies of "standard" versus "non-standard" language varieties. Pedagogically, teachers need to contextualize LLM-mediated listening practice within broader exposure to authentic human speech.

Speaking and Interactive Dialogue

Perhaps the most visible application of LLMs in foreign language education is conversational practice, where learners interact with the model as a simulated partner in the target language [6] [9]. Exploratory studies with university students indicate that learners perceive ChatGPT-like systems as useful for developing speaking fluency, reducing anxiety, and experimenting with new vocabulary and grammatical structures in a low-stakes environment [9] [10]. The model can role-play different interlocutors (e.g., a travel agent, colleague, or examiner), adjust complexity on demand, and maintain topic coherence across multi-turn dialogues.

However, current LLMs exhibit notable limitations as speaking tutors. They cannot fully simulate turn-taking dynamics, non-verbal cues, or socio-pragmatic nuances of real-time human interaction, and their feedback on spoken output is typically text-based and post hoc [4] [6]. There is also a risk that learners may treat the LLM as an infallible authority, accepting its corrections without reflection and thereby weakening metalinguistic awareness. Recent work argues that LLMs can be effective speaking tutors only when learners are guided to critically evaluate feedback, compare alternative formulations, and reflect on communicative appropriateness rather than merely mimicking the model's output [6].

Reading and Vocabulary Development

LLMs can support reading skills by generating or adapting texts to specific proficiency levels, genres, and topics, as well as by providing instant glosses, paraphrases, and comprehension questions [1] [2]. Studies on AI-mediated reading suggest that such scaffolding can enhance learners' engagement and help them access more challenging materials, particularly when combined with strategies such as reading aloud, guided questioning, and vocabulary recycling [3] [10]. Because LLMs can dynamically rephrase passages and create multiple contextualized examples, they are well-suited for deepening lexical knowledge and collocational awareness.

At the same time, heavy reliance on automated simplification and instant translation may undermine the development of independent reading strategies, including inferencing from context and tolerating ambiguity [4] [5]. Moreover, LLMs

occasionally misinterpret domain-specific terminology or cultural references, which can mislead learners if teacher oversight is absent. Effective integration, therefore, requires task designs that balance LLM-provided support with opportunities for learners to grapple productively with authentic, unsimplified texts.

Writing, Feedback, and Revision

Writing is arguably the domain where LLMs have demonstrated the most immediate impact, offering learners on-demand assistance with idea generation, organization, grammar correction, and style refinement [5] [10]. Empirical work with English as a foreign language (EFL) students shows that learners frequently use ChatGPT to draft and revise essays, practice email writing, and receive detailed feedback on errors and discourse structure [10]. Students often report high satisfaction with the immediacy and specificity of this feedback, which contrasts with the delays and limited depth of traditional teacher-provided comments in large classes [10] [11].

Yet, this convenience introduces tensions around authorship, academic integrity, and the depth of learning. Systematic reviews highlight that uncritical use of LLMs for writing can lead to superficial understanding, diminished practice of core skills such as paraphrasing and synthesis, and a tendency to outsource cognitive effort [4] [5]. To harness LLMs constructively, educators are beginning to frame them as “revision partners” rather than ghostwriters, requiring learners to annotate model suggestions, justify accepted changes, and compare multiple generated versions before finalizing their texts [4] [6].

Summary of Pedagogical Affordances

Table 1 summarizes key pedagogical affordances of LLMs across the four language skills, along with illustrative benefits and constraints.

Table 1. Pedagogical affordances of LLMs across language skills.

Skill	Main LLM-Based Affordance	Reported Benefits	Key Constraints and Risks
Listening	Generation of customized audio scripts and tasks	Improved comprehension with tailored materials [3]	Dependence on TTS quality; limited phonetic depth [3]
Speaking	Simulated dialogue and role-play	Reduced anxiety, increased fluency practice [9] [10]	Limited pragmatics, risk of blind trust [4] [6]
Reading	Adaptive texts, glosses, and comprehension support	Greater access to challenging texts [2] [10]	Potential erosion of independent strategies [4] [5]
Writing	Drafting assistance and formative feedback	Faster, more detailed feedback on texts [5] [10]	Academic integrity, shallow processing [4] [5]

4. Learner Attitudes and Differential Effects

Initial empirical studies show generally positive learner attitudes toward LLM-supported language learning, with students emphasizing the accessibility, responsiveness, and non-judgmental nature of conversational agents [9] [10]. Many learners report that LLMs help them practice outside class hours, experiment with new structures, and receive clarification in their first language when needed, thereby

complementing human instruction rather than replacing it [10] [11]. These findings align with broader educational research that frames LLMs as tools for extending learning opportunities beyond traditional classroom boundaries [1] [12].

However, the impact of LLMs appears to be uneven across proficiency levels and learner profiles. A systematic review of LLMs in education indicates that more competent learners tend to engage with these tools strategically—using them to refine ideas, check understanding, and integrate feedback—while less experienced learners are more prone to over-reliance and passive copying [4]. This pattern suggests that LLMs may amplify existing differences in self-regulation and metacognitive skills unless explicit guidance is provided. Consequently, teacher training and curriculum design must address not only technical use but also critical engagement, self-monitoring, and ethical decision-making in LLM-mediated tasks [4] [5].

5. Risks, Ethical Concerns, and Pedagogical Constraints

Cognitive and Pedagogical Risks

A growing body of literature warns that uncritical integration of LLMs in education may reduce deep cognitive engagement, weaken memory consolidation, and foster surface-level learning [4]. When learners offload comprehension, production, and evaluation tasks to an LLM, they may bypass essential processes such as retrieval practice, elaboration, and error-driven revision that underpin durable language acquisition [4] [5]. For foreign language learners, this could translate into a paradoxical outcome: increased short-term performance on assignments but slower long-term development of autonomous communicative competence.

Pedagogically, LLMs can also disrupt traditional assessment practices. If writing tasks are easily outsourced to generative systems, instructors must reconsider how to evaluate learners' abilities, perhaps shifting toward in-class performance, oral examinations, and process-oriented portfolios [5]. At the same time, excessive surveillance or restrictive policies may undermine students' opportunities to learn how to use LLMs responsibly—a skill that is increasingly relevant in academic and professional contexts [11] [12].

Bias, Representation, and Cultural Issues

LLMs inherit biases from their training data, which can manifest in stereotypical depictions of cultures, gendered language, or unequal treatment of different language varieties [4]. In foreign language education, such biases may subtly shape learners' perceptions of the target language community and reinforce hierarchical valuations of “native” norms. Moreover, the dominance of English-centric corpora can lead to uneven performance across target languages, with higher accuracy and fluency in English than in less-represented languages [2] [7].

Mitigation strategies include prompt-level interventions (e.g., instructing the model to adopt inclusive, non-stereotypical language), systematic bias audits of educational prompts, and the integration of critical discussions about representation into language curricula [4] [6]. Importantly, teachers and learners should be

made aware that LLM output is not neutral but reflects particular linguistic and cultural distributions.

Privacy, Data Protection, and Institutional Governance

When learners interact with LLM-based tools, their prompts, drafts, and personal information may be stored and used to train further or refine models, raising privacy and data protection concerns [4] [11]. This is particularly sensitive in educational settings, where minors may be involved and where institutional policies must comply with legal regulations such as GDPR or comparable frameworks. Institutions implementing LLM-mediated language learning need clear governance structures: transparent terms of use, data-minimization practices, and options for local or privacy-preserving deployments [4] [11].

From a pedagogical standpoint, transparency about data use is also essential for maintaining trust. Learners should understand what information is collected, how it is processed, and what rights they have regarding their data. This ethical literacy is part of a broader need to prepare students for participation in AI-mediated societies [4] [12].

6. Design Principles for Responsible Use in Foreign Language Education

Recent conceptual and empirical work points to several design principles that can help align LLM use with sound language-learning pedagogy [2] [4] [6].

First, LLMs should be framed as supplementary tools that extend, rather than replace, human instruction and peer interaction. This implies designing tasks in which the model supports practice (e.g., targeted feedback, additional examples, or alternative phrasings) while key communicative and reflective activities remain learner-driven and socially grounded [4] [6].

Second, educators should embed explicit metacognitive scaffolds around LLM use, such as requiring learners to explain why they accept or reject particular suggestions, to compare model outputs with their own drafts, or to critique the adequacy of generated responses [4] [5]. These practices encourage active processing and help prevent passive dependence. For example, a writing assignment might mandate that students submit both their original text and a marked-up version showing which LLM suggestions they adopted and why.

Third, institutions need clear pedagogical and ethical guidelines that articulate acceptable uses of LLMs in language courses, including expectations for attribution, collaboration, and integrity [5] [11]. Such policies should be developed collaboratively with teachers, students, and administrators to balance innovation with fairness and accountability. Finally, ongoing professional development is crucial: language teachers require opportunities to experiment with LLMs, share practices, and critically examine their own evolving roles in AI-enhanced classrooms [1] [12].

7. Conclusions

LLMs represent a significant technological advance for foreign language education,

offering scalable opportunities for personalized practice, immediate feedback, and immersive interaction across the four core skills [1] [2]. Yet their generative nature introduces complex cognitive, pedagogical, and ethical challenges that preclude naive adoption. The emerging literature suggests that LLMs can function as effective language learning partners when their use is deliberately structured, transparently governed, and embedded within pedagogies that foster critical engagement rather than substitution of human effort [4]-[6].

Future research should move beyond initial attitude surveys and case studies toward rigorous, longitudinal investigations that compare different designs of LLM-integrated tasks, examine differential effects across learner profiles, and explore the long-term impact on autonomous language competence. Such work will be essential for translating the promise of LLMs in foreign language education into sustainable, equitable, and pedagogically grounded practice [4] [7] [9].

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Dawood, H. (2025) Applications of Large Language Models in Education. Hamdan Bin Mohammed Smart University. <https://www.hu.ac.ae/knowledge-update/from-different-corners/applications-of-large-language-models-in-education>
- [2] Qin, Y.X., Zhou, Y.C., Huang, X.H. and Bao, H.C. (2024) Empowering Foreign Language Education with Large Language Models: Applications and Insights. *Journal of Higher Education Teaching*, **2**, 100-101. <http://stemmpress.com/jhet/jhet20251/2048.html>
- [3] Cherednichenko, O., Yanholenko, O., Badan, A., Onishchenko, N. and Akopiants, N. (2024) Large Language Models for Foreign Language Acquisition. *8th International Conference on Computational Linguistics and Intelligent Systems (CoLInS-2024)*, Lviv, 12-13 April 2024. <https://ceur-ws.org/Vol-3722/paper8.pdf>
- [4] Delikoura, I., Fung, Y.R. and Hui, P. (2022) From Superficial Outputs to Superficial Learning: Risks of Large Language Models in Education. <https://arxiv.org/html/2509.21972v1>
- [5] Aljohani, N.J. (2026) ChatGPT in Language Learning: A Systematic Review of Applications and Challenges. *Social Sciences & Humanities Open*, **13**, Article ID: 102357. <https://www.sciencedirect.com/science/article/pii/S2590291125010885>
- [6] Ye, J.H., Wang, S., Zou, D.Q., Yan, Y.B., Wang, K., Zheng, H.-T., *et al.* (2025) Position: LLMs Can Be Good Tutors in Foreign Language Education. arXiv: 2502.05467v1. <https://arxiv.org/html/2502.05467v1>
- [7] Ye, Y.F., Zhang, Z.Y., Ma, T.Y., Wang, Z.H., Li, Y.Y., Hou, S.F., *et al.* (2025) LLMs4All: A Review of Large Language Models Across Academic Disciplines. arXiv: 2509.19580. <https://arxiv.org/abs/2509.19580>
- [8] Xu, L.Q., Li, Q.W., Peng, T.S., Li, Z.C., Zhao, H. and Wang, P. (2025) Can Large Language Models Be Good Language Teachers? *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, Suzhou, 4-9 November 2025, 23957-23971.

- [9] Klimova, B., Pikhart, M. and Al-Obaydi, L.H. (2024) Exploring the Potential of ChatGPT for Foreign Language Education at the University Level. *Frontiers in Psychology*, **15**, Article 1269319. <https://doi.org/10.3389/fpsyg.2024.1269319>
- [10] Al-Obaydi, L.H., Pikhart, M. and Hossain, M.K. (2025) ChatGPT and the Development of Core Language Skills: An Exploratory Study of EFL College Students. *Contemporary Educational Technology*, **17**, ep591. <https://www.cedtech.net/download/chatgpt-and-the-development-of-core-language-skills-an-exploratory-study-of-efl-college-students-17242.pdf>
- [11] Senechal, J., Ekholm, E., Aljudaibi, S., Strawderman, M. and Parthemos, C. (2023) Balancing the Benefits and Risks of AI Large Language Models in K12 Public Schools. VCU Scholars Compass. https://scholarscompass.vcu.edu/cgi/viewcontent.cgi?article=1133&context=merc_pubs
- [12] Saiedian, H. (2024) Leveraging Large Language Models in Education: Enhancing Learning and Teaching. ASEE PEER. <https://peer.asee.org/leveraging-large-language-models-in-education-enhancing-learning-and-teaching.pdf>