Applied Language Learning

Research Article

# Generative AI in Foreign Language Education: A Scoping Review of Pedagogical Applications and Challenges

**Yuliya Kozlova**
*Defense Language Institute Foreign Language Center, Fort Meade Language Detachment, Maryland*

*Generative artificial intelligence (AI) has quickly become an important force in education, especially in foreign language (FL) teaching and learning. While research on generative AI in education is growing, its value and the real challenges of using generative AI in language classrooms are not fully clear. This scoping review explores a dataset of 32 studies on generative AI's role in FL education, published between 2023 and January 2026. Specifically, it addresses: (1) how generative AI tools have been used to support language learning and teaching; (2) what instructional benefits and pedagogical affordances have been reported in relation to generative AI use; and (3) what pedagogical, technical, and ethical issues have been identified in their implementation. Following PRISMA-ScR guidelines, studies were systematically identified, screened, and coded based on their alignment with the research questions. The literature review findings reveal that generative AI tools are used in FL education as tutors that provide personalized feedback, as platforms for simulated communication through chatbots, and as a support for teachers in lesson planning and developing materials. Pedagogical affordances associated with the use of generative AI include increased teaching efficiency, improved learner engagement, more tailored feedback, and broader opportunities for personalized language practice. Finally, among challenges in incorporating generative AI in FL teaching and learning, the most common are uneven teacher preparedness, insufficient support from educational institutions, and concerns about academic integrity and data privacy. The gaps in cumulative knowledge and suggestions for future research are discussed at the end of the scoping review.*

**Keywords**: *Generative AI, Foreign Language Education, Teacher Training, Pedagogy, Artificial Intelligence Literacy, Learner Autonomy, Scoping Review*

## INTRODUCTION

Since the launch of ChatGPT in November 2022, the use of generative artificial intelligence (AI) has expanded rapidly in foreign language (FL) education. Unlike earlier technology tools, which were primarily used to deliver pre-programmed content, generative AI is able to create original language in context. Kasneci et al. (2023) describe this function as an ability that challenges conventional boundaries between learner, teacher, and tool. Research shows that generative AI can lower learners' anxiety when speaking in an FL (Wan & Moorhouse, 2025; Wang et al., 2024; Wang et al., 2025), improve skills development (Cakmak, 2025; Kohnke et al., 2023; Liao et al., 2023), and offer personalized help (Creely, 2024). Additionally, generative AI can foster learners' autonomy and reduce cognitive load (Feng, 2025). In instruction and assessment contexts, generative AI has been explored for the creation of materials and test items (Pack & Maloney, 2023; Perez-Nunez, 2023; Vigna-Taglianti, 2024), feedback (Roe et al., 2024; Zhang & Wang, 2025), and assisting with grading (Barrett & Pack, 2023; Zacharis & Papadakis, 2025). As empirical research on generative AI in FL education continues to grow, attention has shifted toward synthesizing the existing work to clarify predominant trends, benefits, and challenges of integrating this technology into practice.

Recent research has synthesized how generative AI can be incorporated into the FL classroom (Crosthwaite et al., 2025; Luo et al., 2025; Feng et al., 2025) with the prime emphasis on ChatGPT use (Hong, 2023; Li et al., 2024), English as a Foreign Language (EFL) context (Balcı, 2024; Lo et al., 2024), and narrow pedagogical areas of focus (Crosthwaite & Sun, 2025; Du & Daniel, 2024), leaving other languages, educational contexts, AI models, and theoretical perspectives underexplored. In response to this gap in understanding, this literature review takes a scoping review approach to systematically map how generative AI is situated within FL education as a pedagogical tool, including how it may be utilized in lesson planning, developing materials, and assessing learners. The review addresses the following research questions:

1. How have generative AI tools been applied to support foreign language teaching and learning?
2. What instructional benefits and pedagogical affordances have been reported in relation to generative AI use in foreign language teaching and learning?
3. What pedagogical, institutional, and ethical issues and challenges have emerged in the implementation of generative AI in foreign language teaching and learning?

By systematically mapping 32 empirical studies and coding their publication contexts, educational settings, types of generative AI tools, pedagogical applications, and reported benefits and challenges, this scoping review provides an evidence-based overview to inform future research and instructional practice in FL education.

# METHODS

## Operational Definitions

In this review, a foreign language refers to any language that is not the learner's native language and is not commonly used by the learner for daily communication, education, or official purposes (Richards & Schmidt, 2013). This definition does not distinguish between foreign and second language contexts and includes learning both inside and outside the target-language environment. Generative AI is understood as a subset of AI that can create original language outputs, typically through large language models used in tools like ChatGPT and Gemini. More broadly, AI refers to computer systems designed to complete tasks related to human intelligence, including earlier rule-based or automated writing evaluation tools that do not create new language. Large language models (LLMs) are data-driven neural network models trained on extensive textual data to produce human-like language. This study used a scoping review methodology to explore the breadth and nature of research on generative AI in FL education, following PRISMA-ScR guidelines (Tricco et al., 2018). Lastly, gray literature refers to non-peer-reviewed scholarly sources, such as reports or working papers, that were included to gather early insights on the initial stages of generative AI adoption in education.

## Study Design

As generative AI began to gain traction in the research field only towards the end of November 2022, despite having been in development for years, there is insufficient literature to conduct a systematic literature review. This situation makes a scoping review a more appropriate choice. Unlike systematic reviews, scoping reviews are favorable for addressing broader questions, such as "What is the nature of the evidence for this intervention?" or "What is known about this concept?" (Delgado-Chaves, 2025). Given the distinct objectives of this study, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines (Tricco et al., 2018) were adopted to ensure methodological transparency and replicability. Following PRISMA-ScR, the process enabled a comprehensive mapping of generative AI applications, challenges, and pedagogical implications in FL education.

## Eligibility Criteria

The documents were included based on the following criteria: a focus on generative AI-powered tools in FL teaching and learning contexts; involved teachers, students, or both as the target population; no language restrictions; and published between 2023 and January 2026. Eligible publication types included peer-reviewed empirical studies, conceptual or theoretical papers, review articles, and discussion papers. Gray literature was also included when it provided substantive scholarly insight into early applications or conceptual discussions of generative AI in language education.

## Search Strategy

The process involved several steps: selecting databases, defining search terms, applying inclusion and exclusion criteria, and critically analyzing the selected literature. The databases were selected based on their relevance to education, applied linguistics, and educational technology research, as well as their coverage of peer-reviewed and interdisciplinary studies. ERIC and Education Source were used to find education-focused literature, ScienceDirect and JSTOR for foundational research, and Google Scholar to find new studies that the traditional databases do not always index. Major phrases were used to identify relevant studies: "generative AI", "Gen AI", "GAI", "Large Language Models", "LLM", "Chatbots", "Chatbot", "Chat GPT", "ChatGPT", "Gemini", "Copilot", coupled with "foreign language", "FL", "foreign language teaching", "foreign language learning", "foreign language instruction", "foreign language education", "impact on foreign language education", "future of foreign language education", "negative impact on foreign language education", "positive impact on foreign language education", or "integration in foreign language." The search strings were adjusted according to the search engine's requirements.

## Selection of Sources of Evidence and Inclusion Criteria

The selection process included two phases. During the first phase, the researcher screened the titles and abstracts to assess their relevance to the study's topic. In the second phase, the full texts that had passed the previous screening were reviewed, with reasons for exclusion documented. The inclusion criteria were limited to empirical research studies that examined generative AI in foreign language teaching and learning and reported original data. The eligible studies used qualitative, quantitative, or mixed-methods designs and focused on the application of generative AI in foreign language education. When multiple publications reported on overlapping datasets or closely related analyses, the most comprehensive or recent study was included. Only studies focusing on instructional practices involving teachers, instructors, or educational institutions were considered. Studies that focused only on student-created work without any teaching context were excluded, as this review concentrated on how generative AI is integrated and implemented in foreign language education.

The flow diagram, which summarizes each step of the selection process, is presented in Figure 1. During the identification phase, an initial 214 records (100%) were retrieved from the scholarly search engine Google Scholar (170) and these selected databases: ERIC (10), Education Source (15), JSTOR (7), and ScienceDirect (12). No additional records were identified through other sources. All citation data were exported to Microsoft Excel for organization and duplicate removal. After this step, 154 unique records (≈72%) remained.

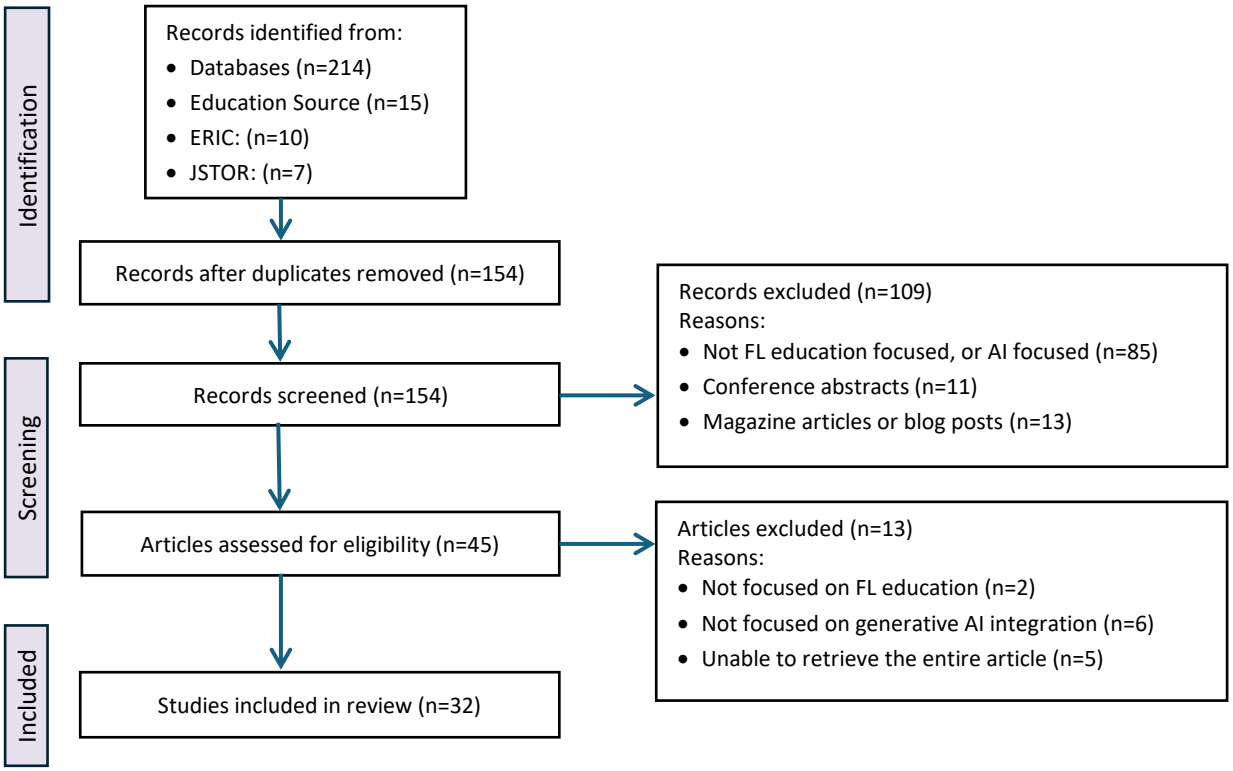
In the screening phase, these records were examined based on titles, abstracts, and relevance to the research focus. A total of 109 records were excluded because they were not related to foreign language education or generative AI, were duplicates of preprints, or originated from non-academic sources such as blogs, conference abstracts, or promotional materials. F For example, studies that offered general overviews of generative AI in FL education without reporting original

empirical data (such as Mizumoto, 2025, which provides a synthesis of prior work rather than a study with clearly defined data sources and analytical methods) were excluded at this stage. The remaining 45 full-text articles (≈21%) were reviewed in detail for eligibility.

Of these, 13 articles were excluded due to insufficient methodological rigor, lack of a clear generative AI focus, or limited alignment with the review questions. For instance, Ulanova et al. (2025) was excluded from the review as it addresses AI broadly rather than specifically exploring generative AI applications. Also, the study does not provide insights into implementing or utilizing these methods in FL teaching and learning. In contrast, Chen (2025) was included because it offers an empirically grounded investigation of how generative AI can be integrated into FL education, with clearly defined research goals and direct relevance to the review questions. Finally, 32 peer-reviewed studies (≈15%) met all inclusion criteria and were included in the final literature review study.

**Figure 1**

*Flowchart Summary of the Sources of Evidence*




## Data Analysis

The data analysis followed an iterative, descriptive-analytical approach consistent with scoping review methodology. After final study selection, all included articles were coded in a Microsoft Excel spreadsheet to enable transparent organization, comparison, and synthesis of findings.

## Coding Process and Trustworthiness

A coding framework was developed deductively based on the study's research questions and refined inductively during initial familiarization with the literature. The final coding scheme was organized around three primary analytical categories aligned with the review questions: pedagogical applications of generative AI (RQ1), reported instructional benefits and affordances (RQ2), and pedagogical, institutional, and ethical challenges (RQ3). Additional contextual categories, including publication information, research design and methodology, and educational context and participant characteristics, were used to describe trends across the literature but were not analyzed as primary outcomes. The categories were organized into separate columns in the Microsoft Excel coding spreadsheet, which allowed a systematic mapping of how generative AI has been understood, used, and assessed in FL education.

Scoping reviews are explanatory and descriptive by nature, aiming to identify and map the literature on a given topic, clarify concepts, and examine how research has been conducted across different settings (Munn et al., 2022). This scoping review follows this approach, presenting the emerging characteristics and patterns of existing research on generative AI integration in foreign language education. Trustworthiness was achieved through qualitative synthesis methods, which included constant comparison across studies and an audit trail in the Excel dataset (Miles et al., 2014; Saldaña, 2011). For example, several publications described how generative AI can be used to enhance writing skills in FL, but the role of AI was not clearly stated, whether it served as support for teachers (e.g., generating prompts, lesson materials, or feedback models), or whether generative AI acted as a learner-facing tool used directly by students. In those cases, the literature was re-examined to determine how generative AI was used in practice and re-coded accordingly. Similarly, when studies reported multiple uses of generative AI in teaching (e.g., both feedback provision and content generation), coding decisions were compared across categories to ensure consistent classification. All revisions were documented in the Excel-based audit trail, which helped to maintain consistency and transparency throughout the review process.

## Analytical Procedures

To analyze data, this scoping review summarized the frequencies and distributions of key study characteristics, such as research designs, types of generative AI applications, and educational contexts. For example, frequency counts were used to summarize study characteristics and were organized into frequency distributions for reporting.

Thematic analysis was used to identify recurring patterns related to benefits, challenges, pedagogical implications, and ethical considerations of generative AI in FL education. During analysis, findings sections of the included studies were examined to extract qualitative evidence describing instructional benefits, constraints, and pedagogical considerations. Themes were developed by repeatedly comparing studies and adjusted to reflect similar and differing viewpoints in the literature, using standard qualitative analysis methods (Braun & Clarke, 2006). The coding categories guiding this analysis and their alignment with the study's review questions

are outlined in Table 1. However, it is important to note that only categories directly aligned with the research questions (categories 4-7) were used in the thematic analysis. The remaining categories (1-3) were used to describe patterns in the literature.

**Table 1**

*Coding Framework and Alignment with Review Questions*

| Number | Coding Category | Description | Related Research Question |
|---|---|---|---|
| 1 | Publication information | Author(s), year of publication, and journal outlet | Contextual descriptor (used to identify trends in the literature; not directly aligned with a specific RQ) |
| 2 | Research design and methodology | Study design, data collection methods, and analytical approaches | Contextual descriptor (used to characterize the evidence base; not directly aligned with a specific RQ) |
| 3 | Educational context and participants | Instructional setting, learner population, and language context | Contextual support for RQ1 (provides insight into where and how generative AI applications occur) |
| 4 | Generative AI tool or model | Type of generative AI system or large language model examined | RQ1: How have generative AI tools been applied to support language teaching and learning? |
| 5 | Pedagogical application | Instructional use (e.g., feedback, writing support, materials development, lesson planning) | RQ1: How have generative AI tools been applied to support language teaching and learning? |
| 6 | Reported benefits and affordances | Positive outcomes and instructional affordances associated with generative AI use | RQ2: What instructional benefits and pedagogical affordances have been reported in relation to generative AI use in foreign language teaching and learning? |
| 7 | Reported challenges and limitations | Pedagogical, technical, ethical, or institutional concerns | RQ3: What pedagogical, institutional, and ethical issues and challenges have emerged in the generative AI implementation in foreign language teaching and learning? |

# RESULTS

## Characteristics of Sources

The review included 32 studies published from 2023 through January 2026. Most studies were quantitative (41%) or mixed-methods (31%), with fewer qualitative designs (28%). The majority were published in 2025, showing rapid growth in this area. Geographically, the studies came from various regions, with China and multi-regional contexts contributing the most.

ChatGPT was the most commonly examined tool at 56%, followed by studies looking at generative AI more generally or using multiple platforms. Most research focused on English as a foreign language, with limited representation of other languages, highlighting an imbalance in the current evidence base (see Table 2 for more details).

**Table 2**

*Frequency of Various Article Characteristics*

| Characteristic Category | Subcategory | Frequency | Percentage |
|---|---|---|---|
| Publication Year | 2023 | 2 | 6.3% |
| | 2024 | 12 | 37.5% |
| | 2025 | 17 | 53.1% |
| | 2026 | 1 | 3.1% |
| Publication Type | Peer-reviewed journal/conference publications | 32 | 100% |
| Study Type | Mixed Methods | 10 | 31.3% |
| | Qualitative | 9 | 28.1% |
| | Quantitative | 13 | 40.6% |
| Country/ Region of Origin | China | 6 | 18.8% |
| | Ukraine | 3 | 9.4% |
| | Russia | 3 | 9.4% |
| | Hong Kong | 2 | 6.3% |
| | Indonesia | 1 | 3.1% |
| | Saudi Arabia | 1 | 3.1% |
| | Poland | 1 | 3.1% |
| | Greece | 1 | 3.1% |
| | Algeria | 1 | 3.1% |
| | Bulgaria | 1 | 3.1% |
| | Colombia | 1 | 3.1% |
| | Italy | 1 | 3.1% |
| | UK | 1 | 3.1% |
| | Turkey | 1 | 3.1% |
| | United States | 1 | 3.1% |
| | Vietnam | 1 | 3.1% |

| Characteristic Category | Subcategory | Frequency | Percentage |
|---|---|---|---|
| | Multi-regional/global* | 6 | 18.8% |
| Generative AI Tool | ChatGPT (explicitly primary tool) | 18 | 56.3% |
| | General Generative AI (unspecified category) | 6 | 18.8% |
| | Multiple AI Tools (ChatGPT + others) | 6 | 18.8% |
| | DeepSeek | 1 | 3.1% |
| | AI Image Generators (text-to-image focus) | 1 | 3.1% |
| FL Focus | English as an FL | 19 | 59.4% |
| | Chinese as an FL | 3 | 9.4% |
| | German as an FL | 3 | 9.4% |
| | Multiple FLs | 5 | 15.6% |
| | French & English (Teacher Education context) | 1 | 3.1% |
| | Unspecified FL (general FL education) | 1 | 3.1% |

*Note.* Three studies were classified as "global" because they analyzed data or synthesized findings across multiple countries rather than focusing on a single national context.

## Results of Individual Sources

The data collection methods, participant types, and research tools used in each publication were examined to provide an overview of the studies included in the final report. The review of 32 studies showed that most research on generative AI in FL education had been conducted in higher education settings, mainly involving university students and teachers. The initial analysis revealed that surveys, mixed-methods, and interviews were the most common instruments, which reflects an increasing interest in learners' perceptions of using generative AI in language learning and its classroom applications. A summary of the study characteristics is available in Appendix A. Also, the detailed findings and pedagogical applications are discussed in the thematic analysis section.

## Results of Thematic Analysis

This scoping review identified the following overarching themes aligned with the research questions: (a) patterns of generative AI use in foreign language education, (b) instructional benefits and learning outcomes, and (c) pedagogical, technical, and institutional issues and challenges.

### Review Question 1: How have generative AI tools been applied to support language teaching and learning?

Analysis of the 32 included studies revealed that generative AI is most commonly used to support writing-related tasks, followed by self-directed learning, instructional design, and conversational practice (see Table 3). Across studies, generative AI functioned primarily as a teacher-supported tool. Writing feedback and revision represented the dominant application. Learners used this technology to generate ideas, refine drafts, correct errors, and improve organization.

Conversational and simulation-based uses included dialogue practice and structured interaction tasks, often supporting speaking preparation.

In teacher-focused contexts, generative AI was used for lesson planning, material development, and assessment design. Additionally, several studies highlighted its role in independent learning within their courses, where students used AI tools for vocabulary development, grammar support, and exam preparation.

**Table 3**

*Theme 1 Categories: Generative AI as an Instructional and Design Support Tool*

| Category | Description | Key Findings |
|---|---|---|
| Tool Distribution | Distribution of generative AI platforms studied across the dataset | • ChatGPT (*n*=18) was the most common tool<br>• General AI and multi-tool studies (*n*=6 each) followed<br>• Few studies examined DeepSeek (*n*=1) or Copilot comparisons (*n*=1) |
| Instructional activity types | Types of pedagogical tasks in which generative AI was integrated | • Writing feedback and revision (*n*=18)<br>• Independent/self-directed learning (*n*=11)<br>• Lesson planning and instructional design (n=10)<br>• Conversational/simulation tasks (*n*=9) |
| Cross-study Trends | Key trends in AI integration across different settings | • Primarily used for writing and structured feedback<br>• Less used for simulations and instructional design<br>• Functioned as a teacher-supported (assistive) tool |

*Note.* Here, *n* refers to the number of studies, not the number of participants.

## Review Question 2: What instructional benefits and pedagogical affordances have been reported in relation to generative AI use in foreign language teaching and learning?

Two themes emerged after the careful analysis of the 32 studies, indicating that the positive pedagogical impact of generative AI in foreign language education can be understood through instructional affordances and differential skill gains in language learning. These themes reflect how AI reshapes instructional processes and the measurable learning outcomes that result from its integration. Table 4 presents the categories for Theme 2 and Theme 3, along with their key findings.

**Table 4**

*Second and Third Themes' Categories*

| Theme | Category | Description | Key Findings |
|---|---|---|---|
| Theme 2: Instructional Affordances of Generative AI | Personalization and adaptive feedback | Generative AI supported personalized, self-paced learning | • Can support grammar and vocabulary development<br>• Promotes learner independence<br>• Particularly beneficial for lower-proficiency learners through tailored feedback |
| Theme 2: Instructional Affordances of Generative AI | Engagement and affective impact | Studies examined learner motivation, confidence, anxiety reduction, and behavioral engagement associated with generative AI use | • Increased motivation and reduced anxiety<br>• Greater willingness to practice<br>• Interactive tasks enhanced engagement |
| Theme 2: Instructional Affordances of Generative AI | Instructional efficiency and teacher support | Teachers' benefits of integrating generative AI in daily instructions | • Saved time on planning, feedback, and content creation<br>• Improved workflow efficiency<br>• Required ongoing pedagogical oversight |
| Theme 3: Writing-Dominant Learning Outcomes | Writing and vocabulary development | Measurable language skill improvements related to AI-supported instruction | • Writing quality<br>• Lexical variety<br>• Grammar accuracy<br>• Vocabulary acquisition<br>• Improved performance through AI-supported drafting and revision |
| Theme 3: Writing-Dominant Learning Outcomes | Reading, speaking, and other skills | Examination of non-writing skill outcomes | • Improved reading comprehension with structured AI tasks<br>• Limited or inconsistent gains in speaking, listening, and intercultural competence |

Generative AI supported personalized, self-paced learning through adaptive feedback and iterative revision (Alkolaly et al., 2025; Chen, 2025; Koka & Khan, 2024; Sun et al., 2025). These features particularly helped lower-proficiency learners, providing targeted support in grammar, vocabulary, and writing development (Karatas et al., 2024; Koka & Khan, 2024). Across studies, increases in learner motivation, confidence, and willingness to practice were observed when generative AI tools were integrated into instruction (Chukhno, 2024; Madison, 2025; Sehlaoui, 2024). Interactive tasks, immediate feedback, and low-pressure practice environments led to higher motivation and less anxiety. However, cognitive engagement relied on task design and instructional guidance (Guo, 2025; Michel et al., 2025; Moorhouse et al., 2024).

From a teaching perspective, generative AI improved efficiency by aiding lesson planning, material creation, and feedback processes (Dornburg & Davin, 2025; Evstigneev, 2024; Murcia et al., 2025). Despite these benefits, studies consistently noted the need for teacher oversight to ensure accuracy and proper teaching methods (Alkolaly et al., 2025; Moorhouse et al., 2024). Regarding learning outcomes, the biggest and most consistent improvements were seen in writing and vocabulary development (Dziemianko, 2025; Koka & Khan, 2024; Madison, 2025; Sun et al., 2025). Learners showed gains in organization, lexical diversity, and grammatical accuracy when working on AI-supported drafting and revision.

In contrast, improvements in speaking, listening, and intercultural competence were limited or inconsistent (Chen et al., 2026; Karatas et al., 2024; Vo & Nguyen, 2024). However, higher-level communication skills depend more on human interaction and contextual classroom practice (Michel et al., 2025; Zhang & Dong, 2024).

### Review Question 3: What pedagogical, institutional, and ethical issues and challenges have emerged in the generative AI implementation in foreign language teaching and learning?

Even though the overall tone of research on generative AI integration is positive, the thematic synthesis identified several important pedagogical, institutional, and ethical challenges, which formed the fourth theme of this scoping review: ethical concerns, accuracy limitations, overreliance, and institutional constraints (see Table 5).

**Table 5**

*Fourth Theme Categories*

| Theme | Categories | Description | Key Findings |
|---|---|---|---|
| Theme 4: Ethical, Technical, and Institutional Implementation Challenges | Ethical and academic integrity concerns | Concerns related to plagiarism, authorship ambiguity, originality, and responsible AI use in FL education | • Risks of overreliance and academic dishonesty<br>• Unclear authorship boundaries<br>• Teachers more concerned than students<br>• Need for AI literacy and ethical guidelines |
| Theme 4: Ethical, Technical, and Institutional Implementation Challenges | Accuracy, reliability, and bias | Limitations of AI-generated content and feedback | • Misinformation and incorrect corrections<br>• Outdated methods and cultural bias<br>• Need for teacher validation of AI outputs |
| Theme 4: Ethical, Technical, and Institutional Implementation Challenges | Overreliance and surface-level learning | Risks of overdependence on AI and limited critical thinking | • Limited critical evaluation of AI outputs<br>• Reduced ownership of writing<br>• Hindered deeper engagement without structured tasks |
| Theme 4: Ethical, Technical, and Institutional Implementation Challenges | Institutional preparedness and policy gaps | Structural and systemic barriers to effective AI integration | • Uneven teacher preparedness and limited PD<br>• Lack of clear policies and infrastructure gaps<br>• Need for training, integration models, and institutional support |

Ethical issues, including plagiarism and authorship ambiguity, were among the most frequently reported concerns (Guo, 2025; Karatas et al., 2024; Pei et al., 2025). Students generally had a positive view of AI (Alkolaly et al., 2025; Sehlaoui, 2024), but instructors were more cautious, stressing the need for clear ethical guidelines and for organized AI literacy instruction (Evstigneev, 2024; Moorhouse et al., 2024; Murcia et al., 2025).

Many studies also reported a challenge related to inaccuracies, biased outputs, and unreliable feedback, underscoring the importance of teacher validation (Dominique et al., 2025; Dziemianko, 2025; Tokmakova & Saenko, 2025; Vigna-Taglianti, 2024). The scholars highlighted

the risks of overreliance, in which learners accepted AI-generated responses without critical evaluation, leading to superficial engagement and reduced ownership of learning (Dominique et al., 2025; Guo, 2025; Huang et al., 2025).

Finally, institutional challenges included uneven teacher preparedness, lack of professional development, and unclear policies governing AI use (Evstigneev, 2024; Moorhouse et al., 2024; Suardika, 2024). Additionally, teacher attitudes toward AI influenced student adoption and classroom integration, stressing the need for coordinated institutional support (Alkolaly et al., 2025; Sun et al., 2025). The literature suggests that successful adoption relies on institutional frameworks that support training, ethical policy development, and effective integration models.

## DISCUSSION

### Significance of the Findings

A synthesis of 32 empirical studies published between January 2023 and January 2026 shows that generative AI offers multiple advantages for FL education. Enhanced personalization, increased learner engagement, and improved instructional efficiency were the most frequent benefits reported across the reviewed studies. Generative AI features, including adaptive feedback, iterative revision, and self-paced practice, supported differentiated instruction and were especially helpful for lower-proficiency learners (Alkolaly et al., 2025; Karatas et al., 2024; Pei et al., 2025; Sun et al., 2025).

The writing assessment showed the strongest and long-lasting learning achievement across all skills. The use of generative AI to assist students with drafting and revision activities resulted in improved organization, lexical diversity, grammatical accuracy, and vocabulary development (Athanassopoulos et al., 2023; Guo, 2025; Huang et al., 2025; Nykyporets et al., 2025). The integration of structured reading tasks improved students' reading proficiency and enhanced their motivation to study the language (Shang et al., 2025). However, generative AI's effects on spontaneous speaking, listening, and intercultural competence were limited or inconsistent (Chen et al., 2026; Vo & Nguyen, 2024).

The literature review also indicates shifts in instructional practice. Teachers who used generative AI to assist them with lesson planning and feedback reported increased efficiency (Dornburg & Davin, 2025; Evstigneev, 2024). However, successful implementation required active teacher involvement and careful assessment of AI outputs (Michel et al., 2025; Zhang & Dong, 2024).

At the same time, recurring concerns related to academic integrity, reliability, overreliance, and institutional preparedness underscore the need for structured integration and clear policy guidance (Chukhno, 2024; Rudnik et al., 2024; Sun et al., 2025). Mainly, the scholars agree that generative AI can support writing development, personalization, and instructional efficiency when thoughtfully integrated into language teaching practice.

## Implications for Research

The findings of this scoping review reveal five directions for future research on generative AI in FL education. The literature published between 2023 and January 2026 shows rapid growth, but the studies remain uneven in scope, method quality, and theoretical foundation.

First, many studies focus on short-term or exploratory outcomes, underscoring the need for longer-term research examining generative AI's lasting effects on language skills, learner independence, and motivation. A significant portion of the existing research has concentrated on immediate outcomes, such as quiz scores and participants' perceived engagement. However, there is a noticeable gap in studies exploring long-term retention and knowledge transfer across contexts.

Second, research must extend beyond outcomes focused mainly on writing. Improvements in writing were consistent, but there is limited evidence for spontaneous speaking, listening comprehension, intercultural skills, and pragmatic development. This gap may exist because writing tasks are easier to design, implement, and assess in controlled research settings. In contrast, speaking and interaction skills are harder to measure and require real-time, dynamic environments. Also, many generative AI tools are currently better suited for text-based output, which may bias research toward writing-focused outcomes. Studies should examine how AI-supported dialogue and real-time interaction can improve communication skills in real-life situations.

Third, research is uneven across languages and regions. Most studies focus on English as an FL and are primarily located in Asia and Eastern Europe, leaving questions about how generative AI impacts less commonly taught languages or multilingual learning settings.

Fourth, more attention should be given to cognitive engagement and AI literacy. Although risks of overreliance were frequently noted, few studies directly measured metacognitive development or critical evaluation skills. Future work should define AI literacy in FL education and explore how structured training and reflective tasks affect learning outcomes. For example, researchers could create interventions that include prompt evaluation, output comparison, and guided reflection, and assess their impact using performance tasks and metacognitive measures. Lastly, research should address teacher professional development and institutional frameworks. Empirical studies on the design of teacher training, policy implementation, and ethical administration remain limited. Comparisons and studies across different contexts would improve the field's theoretical and practical foundations.

## Implications for Practice

The findings of this review suggest that generative AI can be effectively integrated into FL instruction through strategic educational applications. Across studies, the strongest and most consistent benefits were reported in structured writing contexts. Instructors may find it most

effective to use generative AI to develop students' writing skills, with learners using AI to draft their papers, revise them, and analyze provided feedback. However, the evidence also shows that unstructured generative AI use often leads to surface-level revisions that may or may not be necessary or accurate. Writing tasks should include guided comparisons, reflections on AI suggestions, and clear reasons for changes to keep learners cognitively engaged.

At the same time, the evidence does not support replacing face-to-face interaction with AI-mediated dialogue. The studies show that chatbot simulations increased students' willingness to practice and reduced anxiety in low-pressure settings. However, the improvements in spontaneous speaking were limited. In practice, generative AI tools were most effective during rehearsals and useful for preparation and generating ideas. Live communication and instructor feedback remain essential for higher-order proficiency.

Generative AI also supports differentiated instruction through adaptive feedback and self-paced practice. Teachers can use AI to provide targeted grammar clarification and contextualized vocabulary support. Tasks that ask students to evaluate AI outputs, spot inaccuracies, or compare different responses can support the development of critical thinking skills.

Given ongoing concerns about academic integrity and authorship, clear classroom guidelines and institutional policies are essential. Most scholars suggest that banning AI tools might not be a meaningful solution. It is recommended that educators focus on promoting AI literacy and teaching students how to use these tools appropriately in their coursework and assessments.

Finally, effective implementation depends on teacher preparation and institutional alignment. The research shows that the successful use of generative AI depends heavily on how well instructors understand these tools' capabilities and role in FL education. Professional development should go beyond simply showing how to use generative AI. It should focus on designing tasks with clear learning objectives in mind, framing prompts, evaluating AI-generated content, and addressing inaccuracies and bias. Institutional support, including clear guidelines, access to infrastructure, and open discussions across departments, can help reduce uncertainty and encourage effective implementation.

## Limitations

Multiple limitations should be considered when interpreting this scoping review. Although 32 empirical studies published between 2023 and January 2026 were analyzed, the field remains at an early stage. Many studies used short-term interventions, survey designs, or relatively small samples, which limit conclusions about long-term language development and instructional impact. The research base is also uneven in scope. Most studies focused on English as an FL and were conducted mainly in Asia or Europe, with limited representation from less commonly taught languages and other countries. This focus restricts the generalizability of the findings across different learner populations and teaching environments. In addition, the selection of databases used for this review may have limited the inclusion of relevant studies, as relevant research

published in other databases, regional journals, or non-English languages might not have been considered.

The rapid development of generative AI technologies presents an additional challenge. Several studies examined specific tool versions that have changed considerably (e.g., ChatGPT 3, ChatGPT 3.5). As these systems continue to evolve, findings from earlier versions may no longer accurately reflect current capabilities. Methodological diversity makes synthesis more complex. The studies reviewed used different designs, measures, and analytical methods, which made direct comparisons difficult. Lastly, the review included only English-language publications, potentially omitting important research from other languages and regions. Taken together, these limitations show that the findings provide a structured overview of a developing area rather than conclusive evidence of effectiveness. More broad, methodologically rigorous, and diverse research is needed to strengthen the evidence base for generative AI in FL education.

## CONCLUSION

This scoping review of 32 empirical studies suggests that generative AI is most useful in FL education when it is used as a teacher-supported tool, particularly for writing and instructional support. Most studies have linked the use of generative AI to improvements in writing, but its effects on speaking and listening are limited or context-dependent (Athanassopoulos et al., 2023; Guo, 2025; Huang et al., 2025; Chen et al., 2026). At the same time, the findings highlight that generative AI cannot be used effectively without teacher support, as concerns about accuracy, overreliance, and academic integrity remain (Dziemianko, 2025; Murcia et al., 2025; Rudnik et al., 2024). Future research should continue to explore its use across different languages and learning contexts, examine its long-term impact, and identify ways to address the challenges associated with this technology use.

## Author

**Yuliya Kozlova, Ed.D.**, is an Associate Professor and serves as a Cross-Training Branch Chief in the Continuing Education Language, Regional Expertise, and Culture Extension Programs, Defense Language Institute Foreign Language Center. Her areas of interest include the integration of educational technology in foreign language instruction, faculty development, and technology-enhanced learning.

## REFERENCES

Alkolaly, M., Zeid, F., Al-Shamali, N., Khasawneh, M., & Tashtoush, M. (2025). Comparing lecturers and students attitude towards the role of generative artificial intelligence systems in foreign language teaching and learning. *Qubahan Academic Journal*, *5*(3), 1-15. https://doi.org/10.48161/qaj.v5n3a1583

Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural

classroom. *Advances in Mobile Learning Educational Research*, *3*(2), 818-824. https://doi.org/10.25082/AMLER.2023.02.009

Balcı, Ö. (2024). The role of ChatGPT in English as a foreign language (EFL) learning and teaching: A systematic review. *International Journal of Current Educational Studies*, *3*(1), 66-82. https://doi.org/10.46328/ijces.107

Barrett, A., & Pack, A. (2023). Not quite eye to AI: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education, 20*(1), Article 59*. https://doi.org/10.1186/s41239-023-00427-0*

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Cakmak, F. (2025). Generative AI for learning English as a foreign language (EFL). In *AI applications for English language learning* (pp. 293-312). IGI Global Scientific Publishing. doi:10.4018/979-8-3693-9077-1.ch012

Chan, C. K. Y., & Tsi, L. H. (2024). Will generative AI replace teachers in higher education? A study of teacher and student perceptions. *Studies in Educational Evaluation*, *83*, Article 101395. https://doi.org/10.1016/j.stueduc.2024.101395

Chen, M. (2025, May). Generative AI in foreign language inquiry-based learning: A qualitative study of student experiences. In *Proceedings of the 2025 2nd International Conference on Digital Society and Artificial Intelligence* (pp. 151-158). https://doi.org/10.1145/3748825.3748851

Chen, Z., Wei, W., & Zou, D. (2025). Generative AI technology and language learning: Global language learners' responses to ChatGPT videos in social media. *Interactive Learning Environments*, 1-14. https://doi.org/10.1080/10494820.2025.2511248

Chen, X., Zhou, X., & Goh, Y. S. (2026). How Chinese as a foreign language learners use generative AI for oral script-writing: A qualitative perspective on cognitive scaffolding in project-based learning. *Acta Psychologica*, *262*, N.PAG. https://doi.org/10.1016/j.actpsy.2025.106071

Chukhno, O. (2024). Teachers and learners' perspectives on the use of generative AI in foreign language education. *Modern Information Technologies and Innovation Methodologies of Education in Professional Training: Methodology, Theory, Experience, Problems, 73*, 53–60. doi:10.31652/2412-1142-2024-73-53-60

Creely, E. (2024). Exploring the role of generative AI in enhancing language learning: Opportunities and challenges. *International Journal of Changes in Education*, *1*(3), 158-167. https://doi.org/10.47852/bonviewIJCE42022495

Crosthwaite, P., & Sun, S. (2025). Generative AI and L2 written feedback studies: A scoping review. *RELC Journal*, *57*(1). https://doi.org/10.1177/00336882251386530

Delgado-Chaves, F. M., Jennings, M. J., Atalaia, A., Wolff, J., Horvath, R., Mamdouh, Z. M., ... & Baumbach, L. (2025). Transforming literature screening: The emerging role of large language models in systematic reviews. *Proceedings of the National Academy of Sciences*, *122*(2), e2411962122. https://doi.org/10.1073/pnas.2411962122

Dominique, C. V., Huang, J., Polo-Perez, N., & Zaher, A. (2025). Developing writing skills and feedback in foreign language education with ChatGPT: A multilingual

perspective. *Language Learning in Higher Education, 15*(2), 463-483. https://doi.org/10.1515/cercles-2024-0102

Dornburg, A., & Davin, K. J. (2025). ChatGPT in foreign language lesson plan creation: Trends, variability, and historical biases. *ReCALL*, *37*(3), 332–347. doi:10.1017/S0958344024000272

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, *6*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

Evstigneev, M. N. (2024). Planning a foreign language lesson using generative artificial intelligence technologies. *Tambov University Review. Series: Humanities, 29*(3), 617-634. https://doi.org/10.20310/1810-0201-2024-29-3-617-634

Feng, H., Li, K., & Zhang, L. J. (2025). What does AI bring to second language writing? A systematic review (2014-2024). *Language Learning & Technology*, *29*(1). https://doi.org/10.64152/10125/73629

Feng, L. (2025). Investigating the effects of artificial intelligence-assisted language learning strategies on cognitive load and learning outcomes: a comparative study. *Journal of Educational Computing Research*, *62*(8), 1741-1774. https://doi.org/10.1177/07356331241268349

Guo, X. (2025). Facilitator or thinking inhibitor: Understanding the role of ChatGPT-generated written corrective feedback in language learning. *Interactive Learning Environments*, *33*(5), 3526–3544. https://doi.org/10.1080/10494820.2024.2445177

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, *5*(1). https://doi.org/10.61414/jeti.v5i1.103

Huang, J., Mizumoto, A., & Huang, Y. L. (2025). Comparing peer feedback and generative artificial intelligence feedback in Japanese English as a foreign language speaking context: Impacts on motivation, engagement, and writing self-efficacy. *International Journal of Applied Linguistics*, 1. https://doi.org/10.1111/ijal.70020

Hubbard, P., & Schulze, M. (2025). AI and the future of language teaching. *International Journal of Computer-Assisted Language Learning and Teaching*, *15*(1). https://doi.org/10.4018/IJCALLT.378304

Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, *29*(15), 19343-19366. https://doi.org/10.1007/s10639-024-12574-6

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537-550. https://doi.org/10.1177/00336882231162868

Koka, N. A., & Khan, M. R. (2024). Exploring insights from undergraduates on the use of artificial intelligence (AI) tools to enhance proficiency in tense and aspect in foreign language

instruction. *Pakistan Journal of Life & Social Sciences*, *22*(2), 19633–19656. https://doi.org/10.57239/PJLSS-2024-22.2.001440

Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence*, *7*, 100266. https://doi.org/10.1016/j.caeai.2024.100266

Liao, H., Xiao, H., & Hu, B. (2023). Revolutionizing ESL teaching with generative artificial intelligence—Take ChatGPT as an example. *International Journal of New Developments in Education*, *5*(20), 39-46. doi:10.25236/IJNDE.2023.052008

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, *11*(1), 50. https://doi.org/10.1186/s40561-024-00342-5

Luo, Y., Xia, Y. & Lu, X. (2025). A systematic review of the role of artificial intelligence in second language writing education. *Digital Studies in Language and Literature*, *2*(2), 302-325. https://doi.org/10.1515/dsll-2025-0001

Madison Jr, A. (2025). The impact of generative AI on language learning: Transforming pedagogical approaches and learner experiences in the digital age. *TLEP–International Journal of Multidiscipline*, *2*(4), 90-99. https://tlepub.org/index.php/1/article/view/209

Martynyuk, O. (2024). AI tools in foreign language teaching: Pros and cons of Ukrainian and foreign practices. *Comparative Professional Pedagogy, 14*(2), 65–73. https://doi.org/10.31891/2308-4081/2024-14(2)-7

Michel, M., Bazhutkina, I., Abel, N., & Strobl, C. (2025). Collaborative writing based on generative AI models: Revision and deliberation processes in German as a foreign language. *Journal of Second Language Writing*, *67*, N.PAG. https://doi.org/10.1016/j.jslw.2025.101185

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook.* (3rd ed). Sage Publications.

Mizumoto, A. (2025). Generative AI in foreign language learning and teaching: A summary and insights from our empirical investigations. *Information and Technology in Education and Learning*, *5*(1), Inv-p001. https://doi.org/10.12937/itel.5.1.Inv.p001

Moorhouse, B. L., Wan, Y., Wu, C., Kohnke, L., Ho, T. Y., & Kwong, T. (2024). Developing language teachers' professional generative AI competence: An intervention study in an initial language teacher education course. *System*, *125*, 103399. https://doi.org/10.1016/j.system.2024.103399

Munn, Z., Pollock, D., Khalil, H., Alexander, L., Mclnerney, P., Godfrey, C. M., ... & Tricco, A. C. (2022). What are scoping reviews? Providing a formal definition of scoping reviews as a type of evidence synthesis. *JBI Evidence Synthesis*, *20*(4), 950-952. doi*:* 10.11124/JBIES-21-00483

Murcia, D., Osorio, M. F. J., Calderon, L. F. J., & Pipicano, Y. A. H. (2025). GenAI in language teaching, learning, and assessment: Stakeholders insights from two undergraduate language programs. *Lenguaje, 53*(1S), e20314387-e20314387. https://doi.org/10.25100/lenguaje.v53i1S.14387

Nykyporets, S. S., Melnyk, O. D., Piddubchak, S. Y., Melnyk, M. B., & Krutchenko, O. O. (2025). The efficacy of AI-driven pedagogical strategies for foreign language acquisition in higher education. *Society and National Interests, 2*(10), 24-36. https://doi.org/10.52058/3041-1572-2025-2(10)-24-36

Pack, A., & Maloney, J. (2023). Using generative artificial intelligence for language education research: Insights from using OpenAI's ChatGPT. *TESOL Quarterly*, *57*(4), 1571-1582. https://doi.org/10.1002/tesq.3253

Perez-Nunez, A. (2023). Exploring the potential of generative AI (ChatGPT) for foreign language instruction: Applications and challenges. *Hispania*, *106*(3), 355-362. https://dx.doi.org/10.1353/hpn.2023.a906568.

Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching and applied linguistics*. Routledge. https://doi.org/10.4324/9781315833835

Rudnik, Y., Petryk, L., Kosharna, N., Dzhurylo, A., & Popova, L. (2024). The students' awareness considering the academic integrity of artificial intelligence use in terms of foreign language acquisition. *Cadernos de Educação Tecnologia e Sociedade, 17*(3), 50-60. https://doi.org/10.14571/brajets.v17.nse3.50-60

Saldaña, J. (2011). *Fundamentals of qualitative research*. Oxford University Press.

Sehlaoui, F. Z. (2024). Integrating AI in foreign language teaching and learning: Learner autonomy and tool utilization in an Algerian university. *Passerelle, 13*(2), 116-139. https://revue.univ-oran2.dz/Revue/Passerelle/index.php/Passerelle/article/view/361

Shang, J., Huang, Y., Xu, M., Huang, Y., Shen, X., Wang, G., Wang, Y., & Zhang, L. (2025). Competition between human learners and ChatGPT: enhancing university EFL students' reading comprehension and critical thinking through competitive questioning. *Computer Assisted Language Learning*, 1–35. https://doi.org/10.1080/09588221.2025.2577356

Sofronieva, E., Beleva, C., & Georgieva, G. (2025). Use of artificial intelligence in foreign language teaching. *Pedagogy/Pedagogika,* 3961-3982. https://doi.org/10.53656/ped2025-1s.03

Suardika, I. P. G. (2024). Exploring the application of generative artificial intelligence (Gen-AI) at EFL primary class in Kuta District. *Indonesian Journal of Foreign Language Studies*, *1*(1), 10-15. https://doi.org/10.65576/indofes.v1i1.3

Sun, J., Wang, Y., & Qian, Z. (2025). Examining the influence of individual-level cultural values on CFL learners' acceptance of ChatGPT for Chinese learning. *Interactive Learning Environments*, *33*(5), 3393–3407. https://doi.org/10.1080/10494820.2024.2443785

Tokmakova, Y. V., & Saenko, E. S. (2025). The use of corrective feedback from generative artificial intelligence in teaching a professional foreign language to students of an agricultural university. *Tambov University Review. Series: Humanities*, *30*(1), 50-66. https://doi.org/10.20310/1810-0201-2025-30-1-50-66

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467-473. https://doi.org/10.7326/M18-0850

Ulanova, A. E., Buniakova, A. V., & Lazko, A. Y. (2025). Artificial intelligence and foreign language learning in higher education: Analysis of research methods. *Professional Discourse & Communication, 7*(3), 54–68. https://doi.org/10.24833/2687-0126-2025-7-3-54-68

Vigna-Taglianti, J. (2024). AI-Generated images as a teaching tool in foreign language acquisition. *Technology and language*, *5*(3), 85–105. doi:10.48417/technolang.2024.03.07

Vo, A., & Nguyen, H. (2024). Generative artificial intelligence and ChatGPT in language learning: EFL students' perceptions of technology acceptance. *Journal of University Teaching and Learning Practice*, *21*(6), 199-218. doi/10.3316/informit.T202409290000404091943901204

Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, *127*, 103533. https://doi.org/10.1016/j.system.2024.103533

Wang, C., Li, X., & Zou, B. (2025). Revisiting integrated model of technology acceptance among the generative AI-Powered foreign language speaking practice: Through the lens of positive psychology and intrinsic motivation. *European Journal of Education*, *60*(1), e70054. https://doi.org/10.1111/ejed.70054

Wan, Y., & Moorhouse, B. L. (2025). Challenges of AI in language education. *Insights into AI and language teaching and learning*, 32-52. doi:10.29140/9781763711600-03

Young, J. C., & Shishido, M. (2023). Investigating OpenAI's ChatGPT potentials in generating chatbot's dialogue for English as a foreign language learning. *International journal of advanced computer science and applications*, *14*(6). https://doi.org/10.14569/IJACSA.2023.0140607

Zacharis, G., & Papadakis, S. (2025). Can AI grade like a human? Validity, reliability, and fairness in university coursework assessment. *Educational Process: International Journal*, *19*, e2025591. https://doi.org/10.22521/edupij.2025.19.591

Zhang, J., & Wang, J. (2025). Student perceptions of hybrid feedback: Using gen-AI to enhance engagement with EAP writing feedback. *The JALT CALL Journal*, *21*(2), 2175-2175. https://doi.org/10.29140/jaltcall.v21n2.2175

Zhang, Y., & Dong, C. (2024). Exploring the digital transformation of generative AI-assisted foreign language education: A socio-technical systems perspective based on mixed-methods. *Systems*, *12*(11), 462. https://doi.org/10.3390/systems12110462

Zhang, Y., & Miao, M. (2025). The application of generative artificial intelligence in POA guided foreign language teaching. In *Proceedings of the 2024 4th International Conference on Education, Language and Art (ICELA 2024)* (Vol. 907, p. 442). Springer Nature. 10.2991/978-2-38476-364-1_55

## Appendix A

*An Overview of the Studies and Findings of Research Included in the Analysis*

| Author | Year | Educational Context | Type of Generative AI Tool | Research Methods and Participants | Findings |
|---|---|---|---|---|---|
| 1. Alkolaly et al. | 2025 | Higher education; FL education (postgraduate level); universities in Saudi Arabia | Generative AI systems (general reference to generative) | Quantitative survey study; 45 FL lecturers and 493 postgraduate FL students | Students showed higher acceptance than instructors; benefits included efficiency and personalization. |
| 2. Athanassopoulos et al. | 2023 | Secondary education (Junior High School); multilingual and multicultural FL classroom in Greece; learners with refugee/migrant background learning German as a foreign language (A2 level) | ChatGPT (GPT-3.5–based generative AI chatbot) | Empirical classroom-based intervention (pilot study); pre–post writing task design; 8 students aged 15; analysis of lexical and syntactic features | Improved lexical diversity and sentence complexity in writing. |
| 3. Chen | 2025 | Higher education; non-English-major undergraduate students in an FL course; university in western China | Generative AI tools (general reference to generative) | Empirical qualitative study; semi-structured interviews; 23 undergraduate students | Enhanced vocabulary, writing, and critical thinking through personalized support. |
| 4. Chen et al. | 2026 | Higher education; Chinese as an FL instruction | Generative AI tools (general reference to generative) | Empirical qualitative design; participants were university-level Chinese as FL learners | Improved writing quality; limited transfer to spontaneous speaking. |
| 5. Chen et al. | 2025 | Informal and self-directed FL learning; global FL learners using with ChatGPT-related language learning content on social media (YouTube) | ChatGPT | Sequential mixed-methods study. Data consisted of 1,088 first-level YouTube comments collected from 19 highly viewed ChatGPT-for-language-learning videos (Jan–Apr 2023). Participants were global FL learners commenting on social media (no demographic data reported) | Positive perceptions; issues with feedback depth and communication quality. |

| Author | Year | Educational Context | Type of Generative AI Tool | Research Methods and Participants | Findings |
|---|---|---|---|---|---|
| 6. Chukhno | 2024 | FL education, English, secondary and higher education; the Ukrainian context shaped by COVID-19 and the challenges of remote learning during wartime | Generative AI tools (general category; includes AI chatbots and content-generating systems) | Empirical quantitative survey study. 683 participants (110 English language teachers, 573 senior school and university students) from multiple Ukrainian regions | Positive attitudes; concerns about ethics, reliability, and overreliance. |
| 7. Dominique et al. | 2025 | Higher education; FL instruction at Durham University (UK) | ChatGPT | Qualitative empirical design. 18 university students with (four Italian, two per other language: Arabic, Chinese, French, German, Japanese, Russian, Spanish) | Increased motivation; concerns about accuracy and authorship. |
| 8. Dziemianko | 2025 | University-level EFL instruction, Poland | ChatGPT-3.5 and Microsoft Copilot, compared with online dictionaries (COBUILD, LDOCE) and Google Dictionary | Experimental, within-subject design, quantitative analysis of learning outcomes; participants were 128 Polish university students enrolled in English language courses | Strong for immediate comprehension; weaker for long-term retention. |
| 9. Dornburg & Davin | 2025 | FL teacher education and instructional design in higher education in U.S. | ChatGPT (GPT-4) | Quantitative design with document analysis. No human participants were involved. The dataset consisted of 50 AI-generated FL lesson plans | Effective lesson drafts; limited cultural depth and interaction. |
| 10. Evstigneev | 2024 | Higher education - FL teacher education (pre-service FL | Generative AI tools, including chatbots (Mistral AI, ChatGPT variants) and multimodal AI tools | Empirical mixed-methods study; 17 undergraduate pre-service FL teachers in FL education | Positive attitudes; outputs required teacher correction. |

| Author | Year | Educational Context | Type of Generative AI Tool | Research Methods and Participants | Findings |
|---|---|---|---|---|---|
| | | teachers); university-level FL methodology course in Russia | (Twee, Kandinsky, Suno) used for lesson planning and instructional design | | |
| 11. Guo | 2025 | University-level German as an FL; undergraduate students in China | ChatGPT | Mixed method design; 50 undergraduate German FL students (aged 18–21) | More feedback generated than teachers; lower accuracy and acceptance. |
| 12. Huang et al. | 2025 | Higher education; EFL instruction in Japanese universities | ChatGPT | Longitudinal mixed-methods design; participants were 147 first- and second-year undergraduate EFL students | Improved engagement and writing support; concerns about dependence. |
| 13. Karatas et al. | 2024 | Higher education (preparatory FL program); university in Turkey; online distance education | ChatGPT (GPT-3.5-based generative AI chatbot) | Empirical qualitative case study; 13 preparatory-level English FL students | Improved writing and motivation; limited impact on speaking/listening. |
| 14. Koka & Khan | 2024 | Multiple educational contexts in FL learning, (higher education and institutional language programs) | ChatGPT | Mixed-methods design. The data were collected from approximately 450 language learners across multiple institutions | Improved vocabulary, fluency, and engagement; risks of overreliance. |
| 15. Madison | 2025 | Foreign/second language learning across multiple settings (universities, community colleges, and private language schools); multilingual global learner population | ChatGPT, GPT-4, and specialized generative AI language learning applications | Empirical mixed-methods study; 450 language learners (ages 18–65) studying English, Spanish, French, or Mandarin | Improved multiple skills; increased engagement and reduced anxiety. |
| 16. Michel et al. | 2025 | Higher education; German as an FL | ChatGPT | Qualitative, classroom-based design; eight participants (4 pairs), aged 18–23 | Supported collaborative writing and deeper revision processes. |

| Author | Year | Educational Context | Type of Generative AI Tool | Research Methods and Participants | Findings |
|---|---|---|---|---|---|
| 17. Moorhouse et al. | 2024 | Initial Language Teacher Education (ILTE); master's program in English Language Teaching (Hong Kong university). | Primarily ChatGPT; limited exposure to image generators (e.g., Hotpot.ai), writing assistants (Quill Bot), and other generative AI tools | Mixed-methods intervention study; 54 pre-service language teachers (M.Ed. students) | Improved teacher readiness and prompt design skills. |
| 18. Murcia et al. | 2025 | Higher education; two undergraduate English Language Teaching programs in Colombia | ChatGPT and similar generative AI platforms used for text generation, feedback, summarization, and assessment support | Qualitative exploratory case study; 195 undergraduate students, 35 professors, and 2 program chairs | Widespread use; concerns about academic integrity and AI literacy. |
| 19. Nykyporets et al. | 2025 | Higher education; FL instruction at a university in Ukraine | AI-driven instructional tools, including NLP-based automated writing feedback systems and AI chatbots | Mixed-method, multi-case design combining experimental and quasi-experimental interventions; Three cohorts of university English FL learners | Improved grammar, fluency, and engagement; need for teacher guidance. |
| 20. Pei et al. | 2025 | Secondary education; Grade 10 EFL writing instruction in a Hong Kong secondary school | ChatGPT | Quasi-experimental design; 99 EFL 10 grade students | Strong writing gains, especially for lower-level learners. |
| 21. Rudnik et al. | 2024 | Higher education; FL education; Ukrainian university context | Generative AI tools (general category, primarily focused on ChatGPT) | Empirical quantitative survey study; 37 undergraduate students enrolled in pedagogical specialties | Positive perceptions; strong ethical concerns and need for policies. |
| 22. Sehlaoui | 2024 | Higher education; FL teaching and learning (French and English); Algerian university (teacher education context) | AI-powered language learning tools, including Duolingo, ChatGPT, DataBot (generative and adaptive AI tools) | Empirical quantitative study; 120 undergraduate students enrolled in French and English teacher education programs | Increased motivation, autonomy, and learning effectiveness. |

| Author | Year | Educational Context | Type of Generative AI Tool | Research Methods and Participants | Findings |
|---|---|---|---|---|---|
| 23. Shang et al. | 2025 | University-level EFL; undergraduate students | ChatGPT | Quasi-experimental design; participants were 87 Chinese university freshmen | Improved reading comprehension and critical thinking. |
| 24. Sofronieva et al. | 2024 | Higher education; FL teacher education and pedagogy programs; Sofia University "St. Kliment Ohridski," Bulgaria | ChatGPT-3.5 | Empirical quantitative study; 134 undergraduate and postgraduate students in language teaching and pedagogy programs | Difficulty distinguishing AI vs human texts; ethical implications. |
| 25. Suardika | 2024 | Primary education; English as an FL (EFL) instruction; Kuta District, Bali, Indonesia | Generative AI tools for text (ChatGPT), image (Stable Diffusion and DALL·E), audio (AIVA, Soundful, and Murf.ai), and video (Invideo AI) creation | Empirical descriptive qualitative study; 40 EFL primary school teachers from 22 schools | Improved engagement; barriers included infrastructure and training. |
| 26. Sun et al. | 2025 | Chinese as an FL; international university students in China learning Chinese as an FL | ChatGPT | Explanatory mixed-methods design; participants were132 Chinese FL students at the university in China | Positive attitudes linked to usefulness and ease of use. |
| 27. Tokmakova & Saenko | 2025 | Higher education; professional FL instruction at Voronezh State Agrarian University, Russia | DeepSeek | Empirical quasi-experimental study; 86 FL learners | Improved grammar, vocabulary, and professional language use. |
| 28. Vigna-Taglianti | 2024 | Higher education; professional English for specific purposes (ESP); Landscape Architecture students at a non-linguistic university in Russia | AI image generators (e.g., Craiyon v3, Freepik Pikaso, Pixlr, DeepAI, Kandinsky 3.0, NeiroPlod, Shedevrum); text-to-image generative AI tools | Empirical design-based study; comparative analysis of seven open-source AI image generators | Visual tools improved vocabulary and engagement. |

| Author | Year | Educational Context | Type of Generative AI Tool | Research Methods and Participants | Findings |
|---|---|---|---|---|---|
| 29. Vo & Nguyen | 2024 | Higher education; English-majored undergraduate students at universities in Vietnam | ChatGPT | Empirical quantitative study; questionnaire survey with 369 English-major students | Useful for reading/writing; concerns about accuracy and misuse. |
| 30. Young & Shishido | 2023 | EFL; learners using a mobile, voice-enabled chatbot application for speaking and dialogue practice | ChatGPT (GPT-3.5) | Quantitative study design; 450 ChatGPT-generated dialogue samples were analyzed. No direct learner participants were involved | Effective for beginner/intermediate dialogue practice. |
| 31. Zhang & Dong | 2024 | Higher education; FL programs in Chinese universities | Generative AI (general category, primarily focused on ChatGPT) | Mixed-method design; participants included FL teachers working in university-level language programs, university students enrolled in FL courses, and institutional stakeholders. | Improved personalization; concerns about bias and policy gaps. |
| 32. Zhang & Miao | 2025 | Higher education (vocational college FL instruction) in China; POA (Production-Oriented Approach) - guided classroom teaching | Generative AI tools, including ChatGPT and KIMI (Chinese generative AI chatbot) | Empirical practice-based study (instructional design implementation with teacher-led classroom application); no student sample size explicitly reported | Improved efficiency and differentiation; required teacher mediation. |