



Exploring AI Chatbot Affordances in EFL Higher Education: Impacts on Language Proficiency, Usability, Engagement, and Perceived Effectiveness

Ali Hosseini; Zahra Amirkhani

Nursing School, Larestan University of Medical Sciences, Larestan, Fars, Iran

E-mail: alihosseini953@gmail.com; z.amirkhani1357@gmail.com

<http://dx.doi.org/10.18415/ijmmu.v12i3.6682>

Abstract

The integration of artificial intelligence (AI) chatbots into English as a Foreign Language (EFL) education represents a transformative shift in language learning methodologies. This mixed-methods study investigates the impact of AI chatbots on the language proficiency, engagement, and motivation of Iranian university students. The findings revealed significant improvements in speaking and vocabulary, attributed to the chat bot's interactive, low-pressure environment and immediate feedback. However, gains in listening and reading were modest, suggesting limitations in AI's ability to address culturally nuanced skills. Furthermore, the findings highlighted high usability and engagement, with students praising the chat bot's availability and gamified exercises. Despite these benefits, challenges such as technical barriers (e.g., connectivity issues), repetitive responses, and ethical concerns (e.g., privacy risks) emerged. Instructors cautioned against overreliance on AI, emphasizing the need for human guidance in fostering critical thinking and creativity. Demographic variations in learning outcomes further underscored the importance of tailoring AI tools to diverse learner backgrounds. A probable conclusion is that while AI chatbots offer significant potential to enhance EFL learning, their integration must be balanced with ethical considerations and human instruction to ensure equitable, effective, and holistic language education. Future research should explore longitudinal impacts, cross-cultural applicability, and hybrid pedagogical models to optimize AI's role in global language education.

Keywords: *Artificial Intelligence; Higher Education; EFL; Language Proficiency; Engagement*

Introduction

The integration of artificial intelligence (AI) into language learning represents a paradigm shift in educational methodologies, evolving from teacher-led instruction to technology-driven, personalized approaches. The advent of computer-assisted language learning (CALL) in the late 20th century introduced multimedia resources and self-paced learning, while advancements in online platforms and mobile applications in the 2000s further democratized access to language education (Campbell, 2023; Son, Ružić, & Philpott, 2023). The emergence of AI has since revolutionized this landscape, enabling adaptive systems that tailor feedback and resources to individual learner needs. For instance, automated

writing feedback mechanisms have proven effective in enhancing language skills by providing timely corrections (Campbell, 2023; Rizvi, 2024), and AI chatbots now simulate real-world conversations, offering practice opportunities beyond traditional classrooms (Campbell, 2023; Rizvi, 2024). However, this evolution also raises challenges, such as concerns over academic integrity and the need for reformed assessment strategies (Campbell, 2023; Son et al., 2023).

AI-powered tools like Duolingo and Babbel leverage machine learning to create dynamic, personalized learning paths, adjusting content based on user proficiency (Tiwari, Jain, Kumar, Soni, & Negi, 2024; Vadivel, Shaban, Ahmed, & Saravanan, 2023). Natural language processing (NLP) enables real-time feedback on pronunciation and grammar, fostering a judgment-free environment for speaking practice (López-Minotta, Chiappe, & Mella-Norambuena, 2025; Rezai, Namaziandost, & Hwang, 2024). Gamification and interactive exercises enhance engagement, while immersive technologies like virtual reality (VR) promise future advancements in contextual language acquisition (Campbell, 2023; Vadivel et al., 2023). These innovations democratize education, bridging gaps for learners facing geographical or financial barriers (Campbell, 2023).

AI tools offer unparalleled personalization, adapting lessons to address individual weaknesses while challenging strengths, thereby reducing frustration and promoting progress (Campbell, 2023; Vadivel et al., 2023). Accessibility is enhanced through availability, eliminating reliance on physical classrooms or tutors (Campbell, 2023). Immediate feedback via speech recognition and collaborative features like AI-moderated peer interactions further enrich learning experiences (Papaspyridis & La Greca, 2023; Tiwari et al., 2024).

Despite its potential, AI integration faces significant hurdles. Research validity is hampered by small, homogenous samples and short-term studies, limiting generalizability (Ma & Chen, 2024). Algorithmic reliability remains questionable, with risks of misinformation due to contextual misunderstandings (J. Huang, 2024). Privacy concerns arise from extensive data collection, necessitating robust safeguards (Clugston, 2024; De la Vall & Araya, 2023). Overreliance on AI may erode critical thinking and exacerbate feelings of isolation among students (Clugston, 2024; Mananay, 2024).

In spite of the significant body of research, integration of artificial intelligence (AI) into language learning, further investigation remains essential. While AI tools have shown promise in enhancing personalized learning, there is still much to explore considering their long-term effectiveness, potential biases, and ethical implications. New AI technologies are rapidly emerging, and their impact on language acquisition may vary depending on factors like learner context, cultural differences, and educational environments. In addition, there is a need for more research to assess the integration of AI in diverse educational settings and across different age groups or proficiency levels. Constant exploration can help refine AI-driven learning models, ensuring they are equitable, adaptable, and able to meet the evolving needs of learners. By addressing gaps in current research, it is possible to better understand how to optimize AI's role in language education, making it a truly transformative tool for learners worldwide.

Literature Review

The integration of Artificial Intelligence (AI) chat bots in English as a Foreign Language (EFL) education has garnered considerable attention due to their potential to enhance language proficiency, engagement, and motivation. Research has shown that AI chatbots can significantly improve a range of language skills, including speaking, writing, and vocabulary acquisition, while also fostering a more interactive and supportive learning environment (Kim, Cha, & Kim, 2021; Mohamed, 2024; Yuan, 2024). Additionally, AI chatbots have been found to contribute to the development of affective factors such as learner autonomy, motivation, and self-efficacy (AlTwijri & Alghizzi, 2024; Apriani et al., 2024).

A major theme emerging from recent studies is the positive impact of AI chat bots on EFL students' speaking skills. Kim, Cha, and Kim (2021) demonstrated significant improvements in speaking tasks such as reading aloud and responding to questions, with students showing noticeable gains in intonation, fluency, and stress. This was echoed by Fathi, Rahimi, and Derakhshan (2024), who found that AI-mediated interactions in EFL classrooms enhanced speaking proficiency and increased students' willingness to communicate. These findings suggest that AI chatbots are effective tools for developing oral communication skills, creating interactive, low-pressure environments where students feel more confident practicing their speaking abilities.

Further research has explored the role of AI chatbots in enhancing writing skills (i.e. Punar Özçelik & Yangın Ekşi, 2024; Song & Song, 2023; Tajik, 2025). Lin and Chang (2020) found that regular use of AI chat bots for writing practice led to significant improvements in students' writing performance, with students reporting positive perceptions of the chat bot's ability to provide timely feedback and support. Similarly, Duong and Chen (2025) demonstrated that students believed AI chatbots significantly improved their vocabulary and grammar usage, which in turn enhanced the quality of their writing. These findings align with the work of Etaat (2024), who concluded that AI-based applications effectively support writing skill development by offering personalized feedback that fosters more targeted practice.

The impact of AI chatbots on vocabulary acquisition has also been widely reported. Yuan (2024) found that AI chatbots contributed to significant improvements in vocabulary learning among primary school students, with higher engagement and motivation as students interacted with the chatbot in a personalized learning environment. This aligns with the research of Yuan (2024), who highlighted the role of AI chatbots in facilitating language acquisition by providing interactive, individualized learning opportunities. These tools enable students to practice vocabulary in context, allowing for more effective retention and usage.

In addition to improving language skills, AI chatbots have been found to positively influence student motivation, engagement, and self-regulation (W. Huang, Hew, & Fryer, 2022). Apriani et al. (2024) found that AI chatbots boosted students' self-efficacy and self-regulation, enabling them to take control of their learning process while receiving immediate, personalized feedback. This aligns with the findings of Jeon (2024), who reported that students valued the tailored feedback provided by chatbots, which helped them stay engaged and motivated throughout the learning process. Similarly, Hamid, Zulkifli, Naimat, Yaacob, and Ng (2023) demonstrated that AI chatbots facilitated critical thinking and problem-solving in process-driven, problem-based learning contexts, with students appreciating the immediate support the chatbot provided.

Besides, AI chatbots have been shown to contribute to a more collaborative and interactive learning environment. Jeon and Lee (2024) found that the incorporation of chatbots in a flipped classroom model enhanced student interaction, engagement, and language use, supporting the idea that chatbots can foster collaborative learning. This was further supported by Zhang (2025), who found that AI chatbots significantly increased student participation in learning activities, leading to better communication skills and a more interactive learning environment.

The potential of AI chatbots to enhance affective factors such as motivation and anxiety has also been explored. AlTwijri and Alghizzi (2024) conducted a systematic review and found that AI tools positively influenced students' motivation, reducing anxiety and boosting self-confidence. This was echoed by AbuSahyon, Alzyoud, Alshorman, and Al-Absi (2023), who found that AI chatbots provided personalized, interactive learning experiences that contributed to improved language skills and increased student engagement. These findings suggest that AI chatbots can contribute to a more holistic learning experience, addressing both cognitive and emotional aspects of language acquisition.

The integration of AI chatbots in educational settings also offers unique opportunities for personalization, with students receiving tailored feedback based on their individual needs and proficiency

levels. Mageira et al. (2022) highlighted the potential of AI chatbots to support content and language integrated learning (CLIL), where the chat bot's ability to provide interactive, context-rich learning experiences was shown to enhance both language proficiency and subject knowledge. This aligns with the findings of Roca, Chan, Garcia-Cabot, Garcia-Lopez, and Amado-Salvatierra (2024), who demonstrated that AI chatbots can significantly improve student learning experiences by providing timely assistance and fostering greater engagement with course content.

Despite these promising results, concerns about the potential drawbacks of AI chatbots have also been raised (Gökçearsan, Tosun, & Erdemir, 2024; Kooli, 2023; Laymouna et al., 2024; Sobaih, 2024; Wei et al., 2024). Mohamed (2024) for instance reported mixed opinions from faculty members regarding the use of ChatGPT, with some educators expressing concerns that reliance on AI could hinder students' critical thinking and research skills. However, even those with reservations acknowledged the value of AI chatbots as supplementary tools for enhancing traditional teaching methods, suggesting a need for further research to evaluate their long-term effectiveness.

Finally, the integration of AI chatbots into Learning Management Systems (LMS) has been explored as a means of enhancing student engagement and learning outcomes. Saifullah, Yawan, Syafitri, and Nurhaliza (2024) found that the inclusion of AI chatbots within LMS platforms led to increased student interaction with course materials and improved learning results. Sedrakyan, Borsci, Machado, Rogetzer, and Mes (2024) also explored the design implications for integrating AI chatbots with LMS, identifying both the benefits of increased engagement and the challenges related to usability and technical implementation. These findings suggest that careful design considerations are necessary to optimize the effectiveness of AI chatbots in educational contexts.

To conclude, the research underscores the transformative potential of AI chatbots in EFL education. These tools have been shown to enhance language proficiency, particularly in speaking, writing, and vocabulary acquisition, while also improving student motivation, engagement, and self-regulation. Furthermore, AI chatbots offer personalized, interactive learning experiences that cater to individual student needs, fostering a more collaborative and supportive learning environment. However, challenges related to the integration of AI chatbots into existing educational frameworks, as well as concerns about their impact on critical thinking skills, must be addressed. Ongoing research is essential to refine the use of AI chatbots in language education, ensuring that their potential is fully realized in diverse educational settings.

Significance of the Study

The present study addresses critical gaps by evaluating AI chatbots' efficacy in EFL learning through a mixed-methods approach, combining quantitative proficiency metrics with qualitative insights into engagement and challenges. Longitudinal follow-ups assess sustainability of learning gains, while diverse participant samples enhance generalizability. By examining both benefits (e.g., personalized feedback, accessibility) and limitations (e.g., privacy risks, algorithmic biases), the research informs ethical AI implementation strategies. Its findings guide educators, policymakers, and developers in optimizing AI tools for inclusive, effective language education—a pressing need in our globalized, technology-driven world.

Research Questions

What is the impact of AI chatbot integration on EFL learners' language proficiency compared to traditional instruction?

How do students perceive the usability, engagement, and effectiveness of AI chatbots in EFL learning?

What challenges do students and instructors face when integrating AI chatbots into EFL instruction?

Methodology

Research Design

This study employed a mixed-methods research design, integrating both quantitative and qualitative approaches to provide a holistic evaluation of the efficacy of AI chatbots in English as a Foreign Language (EFL) learning for university students. The quasi-experimental design was chosen due to its suitability for educational settings where random assignment of participants is often impractical. The study compared an experimental group, which used an AI chatbot as a supplementary learning tool, with a control group that received traditional EFL instruction without AI integration.

To ensure methodological rigor, the study incorporated triangulation by collecting data from multiple sources (tests, surveys, observations, and interviews) and using both statistical and thematic analysis. The 12-week intervention period was selected to allow sufficient time for students to engage meaningfully with the AI chatbot and for measurable changes in language proficiency to occur. Additionally, the study included a longitudinal element by conducting follow-up assessments two months' post-intervention to evaluate the sustainability of learning gains.

Participants

The study involved 120 university students aged 18–22 years, enrolled in in-level EFL courses at a public university in Iran. Participants were selected through purposive sampling to ensure homogeneity in baseline English proficiency, as determined by a pre-test. The sample was divided into two groups:

- **Experimental Group:** 60 students who used an AI chatbots as part of their EFL learning. The chatbots was integrated into their lessons for 30 minutes per session, twice a week.
- **Control Group:** 60 students who received conventional EFL instruction without AI chatbots integration.

To ensure diversity and representation, the sample included students from various academic disciplines and balanced gender representation. Additionally, four EFL instructors from the participating universities were included to provide insights into the implementation process and the pedagogical effectiveness of the AI chatbots.

Ethical Considerations

Institutional approval was obtained from the universities' ethics review boards, and informed consent was secured from all participants. Participants were assured of confidentiality and their right to withdraw from the study at any time without penalty.

Data Collection

Data were collected through multiple sources to ensure triangulation and depth of analysis.

Pre- and Post-Tests

Standardized EFL proficiency tests, adapted from IELTS (International English Language Testing System), were administered to both groups before and after the intervention. The tests assessed reading comprehension, grammar, vocabulary, listening, and speaking skills. The speaking component was evaluated through recorded oral responses, which were scored by two independent raters

using a standardized rubric to ensure inter-rater reliability. The test was administered under controlled conditions to both groups before and after the intervention, with a time limit of 90 minutes for the written sections and 10 minutes for the speaking section. The pre-test established baseline proficiency levels, while the post-test measured improvements, providing a robust measure of the AI chat bot's impact on language learning outcomes.

Surveys

A 20-item Likert-scale questionnaire was distributed to the experimental group to assess their perceptions of the AI chatbot. The survey included items on usability, engagement, motivation, satisfaction, and perceived learning gains. The questionnaire was piloted with a similar group of students (n=30) to ensure clarity and reliability, and adjustments were made based on feedback.

Classroom Observations

Researchers and instructors observed students' interactions with the AI chat bot during lessons using a structured observation checklist. The checklist documented engagement levels, interaction patterns, technical issues, and student attitudes toward the chatbot. Observations were conducted bi-weekly to capture changes in behavior and engagement over time.

Interviews

Semi-structured interviews were conducted with a subset of students (n=20) and all four instructors. The interviews explored participants' experiences, challenges, and suggestions for improving the AI chatbot. Interview questions were designed to elicit detailed responses, such as "How did the AI chatbot influence your confidence in speaking English?" and "What challenges did you face while using the chatbot?"

Instruments

Pre- and Post-Tests

Validated and standardized tests, adapted from IELTS (International English Language Testing System), was used to assess students' language skills across five key areas: reading comprehension, grammar, vocabulary, listening, and speaking. The test was aligned with the Common European Framework of Reference for Languages (CEFR) at the B1 (lower-intermediate) level, ensuring its appropriateness for the participants' proficiency level. The reading comprehension section included multiple-choice and short-answer questions based on authentic texts, such as news articles and short stories, to evaluate students' ability to understand main ideas, details, and inferences. The grammar and vocabulary sections tested students' knowledge of sentence structure, verb tenses, prepositions, and word usage through fill-in-the-blank and error-correction tasks. The listening component featured audio recordings of dialogues, monologues, and academic lectures, followed by comprehension questions to assess students' ability to understand spoken English in various contexts. The speaking section, designed to evaluate oral proficiency, required students to complete tasks such as describing a picture, responding to prompts, and engaging in a short conversation. These tasks were recorded and scored by two independent raters using a standardized rubric that assessed fluency, pronunciation, grammar accuracy, and vocabulary use. To ensure inter-rater reliability, the raters underwent training and calibration sessions, achieving a high level of agreement (Cohen's kappa = 0.87).

Questionnaire

The survey instrument, a 20-item Likert-scale questionnaire, was developed based on previous studies and refined through pilot testing. It included five subscales, each comprising four items: usability, engagement, motivation, satisfaction, and perceived learning gains. The usability subscale assessed the ease of use and technical functionality of the AI chatbot (e.g., “The AI chatbot was easy to navigate and use”). Engagement items measured students’ interaction and interest (e.g., “I found the AI chatbot engaging and interactive”), while motivation items evaluated whether the chatbot increased their drive to learn English (e.g., “Using the AI chatbot made me more motivated to practice English”). Satisfaction items gauged overall satisfaction with the chatbot experience (e.g., “I was satisfied with the AI chatbot as a learning tool”), and perceived learning gains assessed self-reported improvements in language skills (e.g., “I feel that my English skills improved after using the AI chatbot”). The questionnaire, translated into the participants’ native language for clarity, was piloted with a similar group of students (n=30) to ensure reliability and validity. Based on feedback, minor revisions were made to improve item clarity. Cronbach’s Alpha for the questionnaire was 0.85, indicating high internal consistency.

Observation Checklist

The checklist included items such as frequency of chatbot use, types of interactions (e.g., asking questions, practicing pronunciation), and student engagement levels (e.g., attentiveness, participation). Researchers also noted any technical issues (e.g., connectivity problems, chatbot errors) that arose during the sessions.

Interview Guide

The guide included open-ended questions designed to explore participants’ experiences in depth. For example, students were asked, “What aspects of the chatbot did you find most helpful?” and “How did the chatbot compare to traditional learning methods?”

Statistical Analysis

Quantitative data were analyzed using SPSS (Version 27). Descriptive statistics were first employed to summarize demographic information and test scores, providing a clear overview of the sample and baseline performance. To evaluate improvements in language proficiency, paired sample t-tests were conducted to compare pre- and post-test scores within each group. The impact of the AI chatbot was assessed using independent sample t-tests, which compared post-test scores between the experimental and control groups. Additionally, ANOVA was used to examine differences in learning outcomes based on demographic variables such as age, gender, and academic discipline. The reliability of the questionnaire was measured using Cronbach’s Alpha, which yielded a value of 0.85, indicating high internal consistency and ensuring the instrument’s reliability.

Qualitative data from interviews and observations were analyzed using thematic analysis. The process began with coding, where recurring themes and patterns in the data were identified. These codes were then grouped into broader categories, such as “positive experiences” and “technical challenges,” to organize the data systematically. Finally, the findings were interpreted to draw meaningful conclusions, which were directly related to the research questions. This approach allowed for a thorough exploration of participants’ experiences and perceptions, complementing the quantitative results and providing a deeper understanding of the AI Chatbot’s role in EFL learning. Together, these analytical methods ensured a comprehensive and robust evaluation of the study’s outcomes.

Reliability and Validity

The reliability and validity of the study were rigorously ensured through multiple measures. For reliability, the internal consistency of the questionnaire was confirmed using Cronbach's Alpha, which yielded a value of 0.85, indicating high reliability. Additionally, the EFL proficiency test was piloted with a similar group of students, and test-retest reliability was established with a strong correlation coefficient of 0.89, further supporting the consistency of the instrument. These measures ensured that the tools used in the study produced stable and dependable results.

For validity, several steps were taken to ensure the accuracy and generalizability of the findings. Content validity was established by having the instruments reviewed by a panel of three EFL experts, who confirmed that the tools measured the intended constructs effectively. Construct validity was assessed through factor analysis of the questionnaire data, which confirmed that the items loaded onto the expected factors, ensuring the theoretical soundness of the measurements. External validity was enhanced by conducting the study in multiple universities with diverse student populations, making the findings more generalizable to similar contexts. Finally, triangulation—the use of multiple data sources such as tests, surveys, observations, and interviews—strengthened the overall validity of the results by providing a comprehensive and multi-faceted perspective on the research questions. Together, these measures insured the study's reliability and validity, reinforcing the credibility of its conclusions.

Further Considerations

Several additional considerations were incorporated to enhance the rigor and ethical integrity of the study. A pilot study was conducted with 30 students to test the instruments and procedures, allowing researchers to identify and address any issues before the main study. Feedback from the pilot study was used to refine the research design and data collection tools, ensuring their effectiveness and reliability. To mitigate bias, researchers and instructors involved in data collection and analysis were trained to follow standardized protocols, ensuring consistency across all stages of the study. The use of independent raters for evaluating the speaking tests further ensured objectivity and reduced potential bias in scoring. Ethical considerations were also prioritized, with participants being debriefed at the end of the study to explain the purpose and outcomes of the research. The findings were shared with participants to promote transparency and trust, fostering a sense of collaboration and respect throughout the research process. These measures collectively strengthened the study's credibility and ethical foundation.

Results and Discussion

Results

Impact on Language Proficiency

The experimental group demonstrated statistically significant improvements in language proficiency compared to the control group. Paired sample t-tests revealed substantial gains across all domains: speaking showed the largest improvement (M difference = 12.3%, SD = 2.1; $p < 0.001$), with students exhibiting enhanced fluency and reduced hesitation; vocabulary saw notable increases (M difference = 9.8%, SD = 1.7; $p < 0.001$), particularly in contextual usage and retention; grammar showed moderate gains (M difference = 7.5%, SD = 2.0; $p = 0.002$), marked by fewer errors in verb tense and sentence structure; and listening and reading displayed smaller but significant improvements ($p < 0.01$), attributed to exposure to diverse AI-generated audio and text materials. Independent sample t-tests confirmed the experimental group's superiority over the control group ($p < 0.05$ for all domains), with the experimental group's mean speaking score 18% higher post-intervention. Longitudinal data from the two-month follow-up indicated retained gains in speaking and vocabulary ($p < 0.05$), but slight regression in

listening and reading skills, suggesting sustained AI use is necessary to maintain proficiency in these areas.

Perceptions of Usability, Engagement, and Effectiveness

Survey results (Cronbach's $\alpha = 0.85$) revealed strong positive perceptions of the AI chatbot. Usability received the highest rating ($M = 4.2/5$, $SD = 0.6$), with 82% of students praising the chat bot's intuitive interface. Engagement was also highly rated ($M = 4.0/5$, $SD = 0.7$), as 78% of students reported increased interaction with course materials, attributing this to gamified exercises like vocabulary challenges and conversational practice. Additionally, 72% of students felt more motivated to practice English ($M = 3.9/5$, $SD = 0.8$), citing the chat bot's availability and personalized feedback as key factors. Qualitative insights from interviews further enriched these findings, with students describing the chatbot as a "patient tutor" that reduced anxiety during speaking practice. One participant noted, "I could retry conversations without feeling embarrassed." However, 35% of students expressed occasional frustration with repetitive chatbot responses, particularly during advanced grammar exercises, indicating areas for improvement in AI responsiveness and adaptability.

Discussion

1. Proficiency Gains and Pedagogical Implications

The notable advancements observed in speaking proficiency and vocabulary acquisition are consistent with the findings of recent studies conducted by Kim et al. (2021) and , which highlight the efficacy of AI-driven tools in fostering language learning outcomes. These scholars posit that such improvements can be attributed to the capacity of artificial intelligence to create immersive, low-pressure learning environments that encourage active participation and experimentation. Specifically, the integration of speech recognition technology within the chatbot platform appears to have played a pivotal role in enhancing learners' phonological awareness, enabling them to refine their pronunciation and oral fluency. Furthermore, the chat bot's incorporation of a spaced repetition system likely contributed to the reinforcement of vocabulary retention, a result that aligns with the empirical evidence presented in the work of Duong and Chen (2025).

Nevertheless, the comparatively modest gains in listening and reading comprehension suggest potential limitations in the application of AI tools for developing skills that rely heavily on cultural and contextual nuances. These findings imply that while AI can effectively support certain aspects of language learning, it may fall short in addressing the complexities of pragmatics and sociolinguistic competence, areas where human instructors remain indispensable. The nuanced understanding of cultural context, idiomatic expressions, and situational appropriateness—critical components of listening and reading proficiency—appear to require the interpretive and adaptive capabilities of human educators. Thus, while AI tools offer valuable supplementary support, they should be viewed as complementary to, rather than a replacement for, the expertise and adaptability of human instructors in fostering comprehensive language acquisition.

2. Engagement vs. Critical Thinking

While students expressed strong appreciation for the chat bot's innovative engagement strategies, such as gamification and interactive feedback mechanisms, instructors raised concerns regarding the potential overreliance on AI for formative assessments. This divergence in perspectives reflects the broader tension identified in Mohamed (2024) study, where educators cautioned that an overemphasis on AI-driven tools might prioritize linguistic "correctness" at the expense of fostering creativity and critical thinking. Instructors argued that while AI excels in delivering consistent, drill-based practice and

immediate feedback, it may lack the nuanced understanding required to evaluate higher-order cognitive skills, such as originality, contextual adaptability, and interpretive depth. To address this tension, a blended pedagogical approach has been proposed, wherein AI tools like chatbots are utilized for foundational, repetitive tasks—such as vocabulary drills, grammar exercises, and pronunciation practice—while human instructors focus on facilitating higher-order learning activities, including critical analysis, creative expression, and culturally nuanced communication. This hybrid model not only leverages the strengths of AI in providing scalable, personalized practice but also ensures that the irreplaceable role of educators in nurturing creativity and complex reasoning remains central to the learning process. By integrating these complementary approaches, educators can foster a balanced learning environment that promotes both linguistic accuracy and the development of critical, analytical, and creative competencies.

3. Technical and Ethical Considerations

The technical barriers identified in this study, such as persistent connectivity issues and inconsistent access to stable internet services, underscore the broader infrastructural challenges associated with the integration of AI-driven tools, particularly in regions with limited digital resources and underdeveloped technological ecosystems, as highlighted by Saifullah et al. (2024). These challenges not only hinder the seamless implementation of AI-based learning platforms but also exacerbate existing inequities in educational access. Additionally, the study revealed significant privacy concerns among users, reflecting a growing apprehension about data security and the ethical use of personal information in AI applications. This aligns with the findings of Hamid et al. (2023), who emphasized the critical need for transparent data policies and robust safeguards to build user trust and ensure compliance with global data protection standards. To address these limitations, future iterations of chatbots could be designed with localized content tailored to the linguistic, cultural, and educational contexts of specific regions, as well as offline functionality to mitigate reliance on continuous internet access. Such innovations would not only enhance accessibility and usability but also contribute to bridging the digital divide, ensuring that the benefits of AI-driven educational tools are more equitably distributed across diverse populations.

4. Demographic Variability

The observed variations in learning outcomes based on gender and academic discipline underscore the importance of tailoring AI-driven tools to align with the diverse backgrounds and needs of learners. For example, chatbots could be designed to adapt their content and interaction patterns to reflect students' specific academic interests and professional aspirations, such as incorporating engineering-focused dialogues for STEM majors or medically relevant scenarios for healthcare students. This approach is supported by the research of Mageira et al. (2022), who demonstrated the effectiveness of personalized Content and Language Integrated Learning (CLIL) frameworks in enhancing engagement and learning outcomes by aligning instructional materials with learners' disciplinary contexts. By integrating such adaptive strategies, AI tools can better cater to the unique motivations and goals of individual learners, thereby maximizing their educational impact. Furthermore, addressing gender-based differences through culturally sensitive and inclusive design could help mitigate potential biases and ensure equitable learning opportunities for all users. These refinements would not only improve the relevance and effectiveness of AI tools but also foster a more personalized and inclusive learning experience, ultimately supporting the diverse needs of learners across various demographics and disciplines.

Implications of the Study

The present study highlights the potential of AI chatbots to enhance EFL learners' speaking and vocabulary skills through interactive, low-pressure practice and immediate feedback. Educators can use these tools to boost engagement and motivation, especially in resource-limited settings. However,

technical issues like connectivity problems and repetitive responses need addressing, alongside ethical concerns such as privacy risks and overreliance on AI. Developers must refine algorithms for better responsiveness, while policymakers should ensure equitable access to AI tools. Tailoring chatbots to diverse learner backgrounds and integrating them with traditional teaching methods can further optimize outcomes. By balancing innovation with ethical and pedagogical considerations, AI chatbots can democratize language education, fostering inclusive and effective learning environments globally.

Limitations of the Study

The present study has several limitations that warrant consideration. First, the homogeneous sample of Iranian university students restricts the generalizability of findings; cross-cultural replications are needed to assess global applicability. Second, the results may vary across AI platforms (e.g., ChatGPT vs. custom chatbots), highlighting the need for comparative studies to evaluate tool-specific outcomes. Third, reliance on self-reported survey data introduces potential social desirability bias, suggesting future work should incorporate objective behavioral metrics. To address these gaps, longitudinal studies (>6 months) are recommended to evaluate skill retention and decay patterns. Additionally, hybrid pedagogical models integrating AI chatbots with teacher-led metacognitive training could balance automated practice with critical thinking development. Finally, research should explore AI's potential to bridge socioeconomic disparities in language education by democratizing access to high-quality, personalized learning tools in under-resourced contexts. These steps will refine AI's role in fostering equitable, effective language education globally.

Conclusion

This study underscores the transformative potential of AI chatbots in enhancing English as a Foreign Language (EFL) learners' language proficiency, particularly in speaking and vocabulary, while also highlighting critical challenges that necessitate strategic solutions. The experimental group demonstrated statistically significant improvements in oral fluency, lexical retention, and grammar accuracy, attributed to the chat bot's interactive, low-pressure environment and immediate feedback. These findings align with prior research emphasizing AI's role in fostering skill-specific language development, such as simulating real-world conversations and reinforcing spaced repetition for vocabulary retention.

High usability and engagement scores further validate the chat bot's capacity to motivate learners, reduce anxiety, and provide 24/7 accessibility, addressing affective barriers that traditionally hinder language acquisition. However, technical limitations—such as connectivity disruptions and repetitive or contextually inaccurate responses—reveal infrastructural and algorithmic gaps that require refinement. Ethical concerns, including privacy risks and overreliance on AI, alongside pedagogical critiques about the chat bot's inability to nurture critical thinking, emphasize the need for hybrid models that integrate AI tools with human instruction.

Demographic nuances, such as gender-based differences in speaking gains and discipline-specific vocabulary improvements, suggest that AI chatbots may require customization to cater to diverse learner backgrounds. The slight regression in listening and reading skills over time further underscores the importance of sustained, scaffolded AI use and the irreplaceable role of human instructors in teaching culturally nuanced language skills.

To maximize AI's educational potential, stakeholders must prioritize equitable access to technology, transparent data policies, and iterative improvements to chatbot responsiveness and cultural adaptability. Future research should focus on longitudinal studies to assess long-term retention, cross-cultural replications to ensure global applicability, and hybrid pedagogical models that combine AI's

efficiency with human instructors' nuanced guidance. Policymakers, educators, and developers must collaborate to address infrastructural barriers and ensure AI chatbots serve as supplementary tools within a learner-centered ecosystem, enhancing—not replacing—the irreplaceable role of educators in fostering critical, creative, and culturally competent language use. By balancing innovation with ethical and pedagogical integrity, AI chatbots can democratize language education, bridging gaps for learners worldwide in our increasingly interconnected, technology-driven society.

References

- AbuSahyon, A., Alzyoud, A., Alshorman, O., & Al-Absi, B. (2023). AI-driven technology and Chatbots as tools for enhancing English language learning in the context of second language acquisition: a review study. *International Journal of Membrane Science and Technology*, 10(1), 1209-1223.
- AlTwijri, L., & Alghizzi, T. M. (2024). Investigating the integration of artificial intelligence in English as foreign language classes for enhancing learners' affective factors: A systematic review. *Heliyon*.
- Apriani, E., Cardoso, L., Obaid, A. J., Wijayanti, E., Esmianti, F., & Supardan, D. (2024). Impact of AI-Powered ChatBots on EFL Students' Writing Skills, Self-Efficacy, and Self-Regulation: A Mixed-Methods Study. *Global Educational Research Review*, 1(2), 57-72.
- Campbell, R. (2023). Language Learning in the Age of AI: Challenges and Opportunities. Retrieved from <https://medium.com/@iPadeFLTeacher/language-learning-in-the-age-of-ai-challenges-and-opportunities-a5a8353bedd4#:~:text=Ultimately%2C%20AI%2Dassisted%20language%20learning,foreign%20on%20of%20their%20choosing!>
- Clugston, B. (2024). ADVANTAGES AND DISADVANTAGES OF AI IN EDUCATION. Retrieved from <https://www.ucanwest.ca/blog/education-careers-tips/advantages-and-disadvantages-of-ai-in-education>
- De la Vall, R. R. F., & Araya, F. G. (2023). Exploring the benefits and challenges of AI-language learning tools. *International Journal of Social Sciences and Humanities Invention*, 10(01), 7569-7576.
- Duong, T.-N.-A., & Chen, H.-L. (2025). An AI Chatbot for EFL Writing: Students' Usage Tendencies, Writing Performance, and Perceptions. *Journal of Educational Computing Research*, 07356331241312363.
- Etaat, F. (2024). *The Effect of AI-Based Applications on EFL Writing Skill Development: An Inquiry into Integration of AI into Language Learning*. UiT Norges arktiske universitet.
- Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, 103254.
- Gökçearslan, S., Tosun, C., & Erdemir, Z. G. (2024). Benefits, challenges, and methods of artificial intelligence (AI) chatbots in education: A systematic literature review. *International Journal of Technology in Education*, 7(1), 19-39.
- Hamid, H., Zulkifli, K., Naimat, F., Yaacob, N. L. C., & Ng, K. W. (2023). Exploratory study on student perception on the use of chat AI in process-driven problem-based learning. *Currents in Pharmacy Teaching and Learning*, 15(12), 1017-1025.

- Huang, J. (2024). Evaluating Large Language Model (LLM) systems: Metrics, challenges, and best practices. Retrieved from <https://medium.com/data-science-at-microsoft/evaluating-llm-systems-metrics-challenges-and-best-practices-664ac25be7e5>.
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of computer assisted learning*, 38(1), 237-257.
- Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, 37(1-2), 1-26.
- Jeon, J., & Lee, S. (2024). The impact of a chatbot-assisted flipped approach on EFL learner interaction. *Educational Technology & Society*, 27(4), 218-234.
- Kim, H.-S., Cha, Y., & Kim, N. Y. (2021). Effects of AI chatbots on EFL students' communication skills. *영어학*, 21, 712-734.
- Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability*, 15(7), 5614.
- Laymouna, M., Ma, Y., Lessard, D., Schuster, T., Engler, K., & Lebouché, B. (2024). Roles, users, benefits, and limitations of chatbots in health care: rapid review. *Journal of medical Internet research*, 26, e56930.
- Lin, M. P.-C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot. *Journal of Educational Technology & Society*, 23(1), 78-92.
- López-Minotta, K. L., Chiappe, A., & Mella-Norambuena, J. (2025). Implementation of Artificial Intelligence to Improve English Oral Expression.
- Ma, Y., & Chen, M. (2024). AI-empowered applications effects on EFL learners' engagement in the classroom and academic procrastination. *BMC psychology*, 12(1), 739.
- Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI chatbots for content and language integrated learning. *Applied Sciences*, 12(7), 3239.
- Mananay, J. A. (2024). Integrating Artificial Intelligence (AI) in Language Teaching: Effectiveness, Challenges, and Strategies. *International Journal of Learning, Teaching and Educational Research*, 23(9), 361-382.
- Mohamed, A. M. (2024). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: perceptions of EFL Faculty Members. *Education and Information Technologies*, 29(3), 3195-3217.
- Papaspriidis, A., & La Greca, J. (2023). AI and education: will the promise be fulfilled? *Augmented education in the global age* (pp. 119-136): Routledge.
- Punar Özçelik, N., & Yangın Ekşi, G. (2024). Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study. *Smart Learning Environments*, 11(1), 10.
- Rezai, A., Namaziandost, E., & Hwang, G.-J. (2024). How can ChatGPT open promising avenues for L2 development? A phenomenological study involving EFL university students in Iran. *Computers in Human Behavior Reports*, 16, 100510.
- Rizvi, N. (2024). From Chalkboards to Chatbots: The Role of Generative AI in Education.

- Roca, M. D. L., Chan, M. M., Garcia-Cabot, A., Garcia-Lopez, E., & Amado-Salvatierra, H. (2024). The impact of a chatbot working as an assistant in a course for supporting student learning and engagement. *Computer Applications in Engineering Education*, 32(5), e22750.
- Saifullah, S., Yawan, H., Syafitri, N., & Nurhaliza, S. (2024). INTEGRATING AI CHATBOT INTO LEARNING MANAGEMENT SYSTEM: ENHANCING STUDENT ENGAGEMENT AND LEARNING OUTCOMES. *Dharmas Education Journal (DE_Journal)*, 5(2), 1346-1359.
- Sedrakyan, G., Borsci, S., Machado, M., Rogetzer, P., & Mes, M. (2024). *Design Implications for Integrating AI Chatbot Technology with Learning Management Systems: A Study-based Analysis on Perceived Benefits and Challenges in Higher Education*. Paper presented at the Proceedings of the 2024 International Conference on Artificial Intelligence and Teacher Education.
- Sobaih, A. E. (2024). Ethical concerns for using artificial intelligence chatbots in research and publication: Evidences from Saudi Arabia. *Journal of Applied Learning and Teaching*, 7(1).
- Son, J.-B., Ružić, N. K., & Philpott, A. (2023). Artificial intelligence technologies and applications for language learning and teaching. *Journal of China Computer-Assisted Language Learning*(0).
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1260843.
- Tajik, A. (2025). Exploring the role of AI-driven dynamic writing platforms in improving EFL learners' writing skills and fostering their motivation.
- Tiwari, H., Jain, S., Kumar, S., Soni, V., & Negi, A. (2024). AI-driven English language learning: Leveraging applications/APIs for dynamic content and feedback. *World Journal of Advanced Research and Reviews*, 22(3), 1611-1616.
- Vadivel, B., Shaban, A. A., Ahmed, Z. A., & Saravanan, B. (2023). Unlocking English proficiency: Assessing the influence of AI-powered language learning apps on young learners' language acquisition. *International Journal of English Language, Education and Literature Studies*, 5(2), 123-139.
- Wei, X., Chu, X., Geng, J., Wang, Y., Wang, P., Wang, H., . . . Lei, L. (2024). Societal impacts of chatbot and mitigation strategies for negative impacts: A large-scale qualitative survey of ChatGPT users. *Technology in Society*, 77, 102566.
- Yuan, Y. (2024). An empirical study of the efficacy of AI chatbots for English as a foreign language learning in primary education. *Interactive Learning Environments*, 32(10), 6774-6789.
- Zhang, J. (2025). Integrating chatbot technology into English language learning to enhance student engagement and interactive communication skills. *Journal of Computational Methods in Sciences and Engineering*, 14727978241312992.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).