



Exploring the potential of AI techniques in teaching English as a foreign language: A systematic literature review

Wesam Saad Almehmadi

Department of English Language and Translation, Saudi Electronic University, Jeddah, Saudi Arabia.

Email: w.almehmadi@seu.edu.sa

Abstract

The rapid evolution of artificial intelligence (AI) technology and its integration into different fields, including language teaching, have inspired a growing body of literature. Scholars have particularly examined the integration of AI techniques into the teaching of English as a foreign language (EFL). However, it is becoming more challenging to identify the most suitable and efficient tools for implementation in EFL education because of the massive amount of innovation. Accordingly, in this systematic review, we examine the latest literature on the integration of AI into EFL teaching. The objective of this study is to explore how AI is being incorporated into this field, its impact on enhancing core English skills, and the potential pedagogical implications. A total of 284 articles published between 2019 and 2023 were initially identified from the most popular databases, including ERIC, ScienceDirect, JSTOR, ProQuest, and Scopus. Following pre-established inclusion and exclusion criteria, 13 papers were selected for the final review. The findings of this review highlight the benefits of using different AI techniques such as chatbots, automated writing evaluation, and writing assistance technologies in the instruction of fundamental EFL skills, namely speaking, listening, and writing. This review also provides useful insights and indicates some promising directions regarding the appropriate and effective application of AI in EFL classrooms.

Keywords: Artificial intelligence, Chatbots, ChatGPT, English as a foreign language, Learning, Systematic literature review, Teaching.

Citation | Almehmadi, W. S. (2024). Exploring the potential of AI techniques in teaching English as a foreign language: A systematic literature review. *Asian Journal of Social Sciences and Management Studies*, 11(2), 22-31. 10.20448/ajssms.v11i2.5576

History:

Received: 2 February 2024

Revised: 18 March 2024

Accepted: 8 April 2024

Published: 19 April 2024

Licensed: This work is licensed under a [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/)

Publisher: Asian Online Journal Publishing Group

Funding: This study received no specific financial support.

Institutional Review Board Statement: Not applicable.

Transparency: The author confirms that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.

Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.

Contents

1. Introduction	23
2. Literature Review	23
3. Methodology	24
4. Conclusion	30
References	30

Contribution of this paper to the literature

This study contributes to the existing literature by reviewing the most recent research on the use of Artificial intelligence (AI) techniques in teaching of English as a foreign language (EFL). It explores the current state of AI integration, evaluates the effectiveness of early efforts, and discusses potential pedagogical implications in this domain.

1. Introduction

Digital technology has been used in language learning of all kinds and at all levels for many years, and there is a large literature on topics such as computer-assisted language learning and technology-enhanced language learning (Schmidt & Strasser, 2022). Educators and students have benefited from a wide range of digital tools both in and beyond the classroom. Computer-mediated material can provide input and feedback that is personalized and responsive to each learner's individual needs, such as their learning style, the time they need to learn, and the context in which they learn best. Flexibility, convenience, and choice are key benefits of such material, helping to provide additional practice for learners in all four basic skills—speaking, listening, reading, and writing—with reduced reliance on the expensive resource of face-to-face interaction with a teacher (Kim, Cha, & Kim, 2021). The rapid adoption of the internet across the world has also vastly increased the volume and accessibility of digital materials and methods available for language learning. These early innovations and the ongoing trend of ever-expanding technological possibilities have paved the way for the world in which we live today. However, it is becoming increasingly difficult amid this massive amount of innovation to identify the most appropriate and effective tools to use in EFL education.

The origins of AI are widely believed to lie in the 1950s, when scholars in the United States began working on the creation of machines that could carry out complex tasks, such as playing a game of chess or deciding what to buy on a shopping tour (Cordeschi, 2007). Throughout the second half of the twentieth century, these developments moved out of the industrial and technical spheres into all walks of life, including education. There has been a prevailing focus on the benefits of big data analytics, especially in combination with increasing computational and memory capacity in computers, allowing educators to assess and track student performance, analyze students' behavior in class, and monitor their use of study aids. On the other hand, scholars have warned about the privacy issues that can arise in this context and the risks of reducing learning to a “numbers game, thereby inviting positivistic and behaviouristic responses” (Godwin-Jones, 2017).

AI in education is here to stay, and the formation of interdisciplinary partnerships will be required if we are to bring to fruition the many benefits it promises (Luckin & Cukurova, 2019). New programs, apps, features, and techniques appear almost daily, many of which seem to offer new affordances and learning opportunities in the field of teaching English to speakers of other languages, even if they were designed for entirely different purposes, such as automatic text generation or text editing for journalists, advertisers, and other professionals. There has not yet been time to explore what is on offer, let alone evaluate the appropriateness and effectiveness of the most familiar and popular AI applications or identify any disadvantages or risks that might arise. This study aims to make a small contribution to this overwhelming task by reviewing the most recent literature on AI in EFL education. We explore how AI is currently being integrated into this field, how effective these early efforts have been, and what the possible pedagogical implications might be.

2. Literature Review

2.1. Background of the Topic

There is no single agreed-upon definition of AI because it is a complex and layered phenomenon that is advancing rapidly in a wide range of different fields (Roschelle, Lester, & Fusco, 2020). It can be helpful to think of AI in three different but complementary ways: “as an ambitious leading edge of computing, as a set of specific capabilities that are rapidly advancing, and as a toolkit for synthesizing (and exploring) possible futures [as well as] inspiring new design concepts” (Roschelle et al., 2020). These approaches highlight the novelty of AI and its anticipated importance in our world, but they focus on the theoretical dimension. Thus, they are a good starting point, but not very helpful to teachers and learners looking for immediate benefits from these new technological advances.

The application of some types of AI to the EFL field has occurred rapidly, most obviously in areas such as administration and assessment, where teaching and learning data can be easily gathered, analyzed, and integrated within wider systems (Godwin-Jones, 2017). According to Singh and Hiran (2022) a new phase of engagement has begun, and thanks to the ability of AI to be embedded within the wider, data-rich environment of a school or college, “Teachers and instructors can collaborate with robots in the form of cobots or humanoid robots and chatbots can perform teacher or instructor-like functions.” One recent study noted that the adoption of AI in education is part of a much broader trend in the emergence of the Internet of Things and that, so far, innovations in this field have been “largely driven by intuition and common-sense extrapolations, rather than being solidly underpinned by research-informed models and frameworks” (Bonfield, Salter, Longmuir, Benson, & Adachi, 2020). It always takes time to embed new ideas into education, and teachers are rightly reluctant to subject their students to new, untested approaches.

Interestingly, however, some recent events, such as the global pandemic of 2019–2022, the turbulence caused by climate change, and the spread of military conflict, have accelerated the adoption of AI systems and processes in all aspects and subjects within the educational field (Baidoo-Anu & Ansah, 2023). Analyses of pedagogic interventions introduced during the COVID pandemic have once again highlighted how economic factors, such as unequal access to reliable broadband provision, can seriously affect the impact of all types of online learning, including AI-supported learning (Cullinan, Flannery, Harold, Lyons, & Palcic, 2021). At the same time, neither planners and educators nor students were prepared for the sudden, widespread integration of AI into everyday learning, which suggests that we should be wary of drawing premature conclusions from research carried out during this unusual period. The insights gathered may well be useful, but there were undoubtedly design flaws,

compromises, and omissions in the rush to provide any educational support at all when institutions were forced into lockdown.

Moreover, there are considerable dangers in rushing to implement new practices without properly evaluating early, experimental work. A large-scale review of research on AI in education across the world before the pandemic found that research was heavily skewed toward just a few countries (most notably, the US, China, Taiwan, and Turkey) and dominated by projects in the subject areas of computer science and science, technology, engineering, and mathematics (Zawacki-Richter, Marín, Bond, & Gouverneur, 2019). Furthermore, there was a predominance of quantitative methods, a lack of longitudinal research, and very few implementation or impact studies (Zawacki-Richter et al., 2019). However, some exceptions to these trends can be found. Taguchi (2015) for example, advocates using the traditional approach of pragmatics along with a highly innovative package of online resources to provide an alternative to studying abroad, with consequent cost savings and convenience for students, not to mention benefits to the environment from reduced travel. Similarly, Divekar et al. (2022) point out the underexplored potential of immersion in digital extended-reality contexts, including social interaction with authentic-sounding speakers of the foreign language. Another intriguing study conducted over a whole year explored how five students used virtual reality (VR) headsets and rooms to meet online and review VR products available on the web (Cowie & Alizadeh, 2022). The study acknowledged its selection bias, as the participants were all enthusiastic AI users in their lives outside education, and they identified multiple risks and challenges, such as ethical and health issues with the equipment, access issues due to cost, and the need for extensive educator training in order to align the teacher's own "signature pedagogies" (Cowie & Alizadeh, 2022) with the AI app and master any technical issues that arise.

The literature reviewed above covers a wide but not exhaustive range of approaches to the impact of AI generally and in the EFL classroom specifically. An important limitation on this body of work, however, is that these studies often involved a small number of participants, and the settings varied across different source languages and geographical areas. This means that when variation in the results is observed, it is difficult to establish whether this was caused by some aspect of interaction with AI or other contextual variables that may not have even been measured. All in all, therefore, the somewhat random and fragmented data are difficult to interpret, and the most obvious gap that emerges is the need for a broader, more consistent view with a larger sample size. A large-scale primary research project would be expensive and impractical, however. Therefore, in this study, we undertook a systematic review of 13 highly relevant articles from within the existing literature. This addresses the issues of small sample sizes and disparate groups since it allows comparisons to be drawn and overarching theories to be devised. This, in turn, can provide a firmer basis for future research in this vital and expanding field.

It is evident from this brief summary of the current research that many existing pedagogical approaches can be adapted for use in new AI-supported teaching and learning programs. This sketch also shows that it is time for a more thorough and systematic exploration of the use of AI in EFL. This brings us to the heart of this review, in which we analyze 13 recent articles written by and for EFL specialists about their use of AI over the last few years.

3. Methodology

3.1. Research Instrument

To investigate the various approaches used in implementing AI technology to teach EFL and the impact of AI use on EFL teaching, a systematic literature review (SLR) was conducted. SLR is defined as a systematic process for identifying, organizing, and summarizing existing evidence on a particular topic to answer specific research questions by presenting a comprehensive study plan (Colquhoun et al., 2014; Tawfik et al., 2019). This methodological approach is useful because it provides a clear and concise mapping of the studies published on specific topics, helps identify research gaps, and provides decision-makers with suitable implications and suggestions (Tricco et al., 2018).

Khan, Kunz, Kleijnen, and Antes (2003) five-step method for conducting SLR was used in this study to conduct a comprehensive search and analyze the articles collected from different databases (i.e. ERIC, ScienceDirect, JSTOR, ProQuest, Scopus, and Google Scholar). These five steps include framing the research questions, identifying relevant publications, assessing each study's quality, summarizing the evidence, and interpreting the findings.

3.2. Research Procedures

Step 1: Framing the Research Questions

Our aim was to investigate the potential of AI techniques in teaching EFL and their role in improving core English skills. This systematic review was designed to answer two key research questions:

1. What are the various methods and approaches used in implementing AI technology to teach EFL, and how do these methods contribute to enhancing student learning outcomes?
2. What are the pedagogical implications of integrating AI techniques into EFL instruction?

Step 2: Identifying Relevant Publications

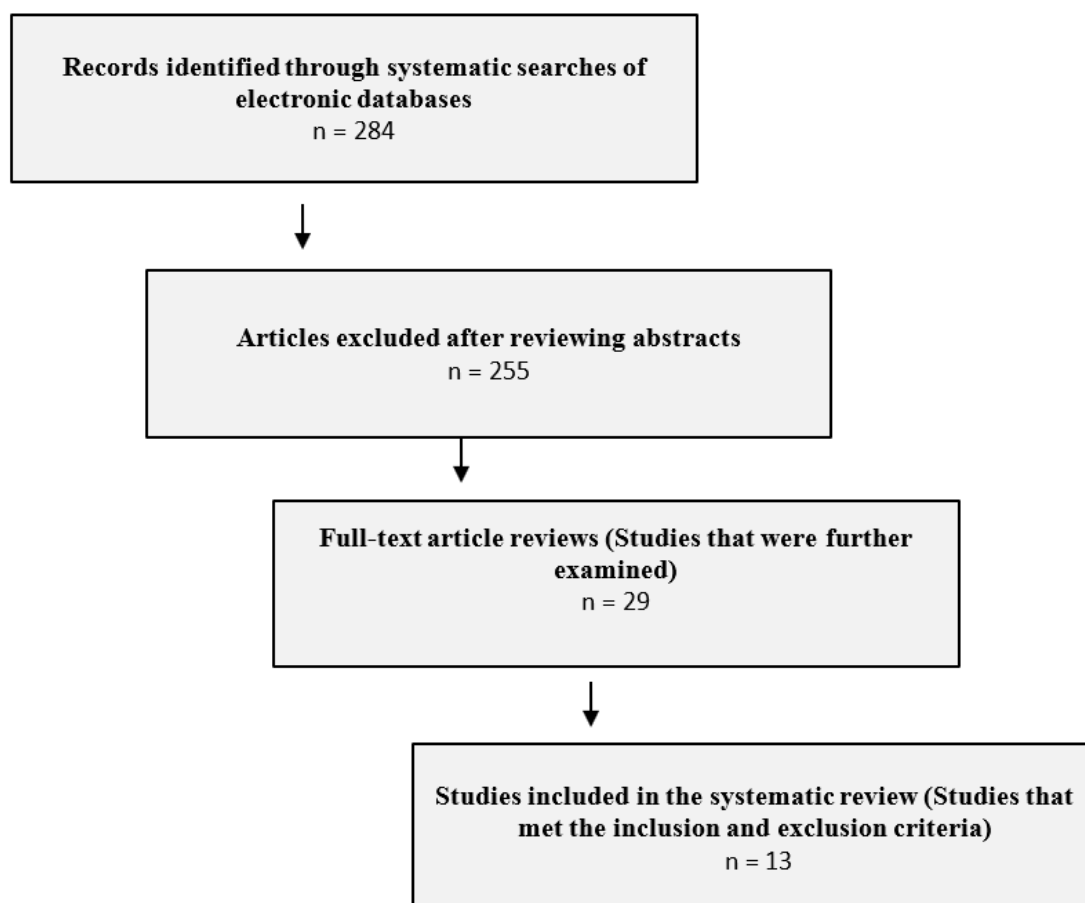
Establishing Inclusion and Exclusion Criteria. This stage involved determining the inclusion and exclusion criteria for selecting relevant sources from the literature. To be included in this review, studies had to meet the selection criteria listed in Table 1. The table also describes the exclusion criteria; any study that did not meet one or more of these criteria was excluded from the review.

Search Strategy. To gather a comprehensive collection of relevant literature, a systematic search strategy was employed that took into account the predetermined criteria for inclusion and exclusion. To capture as many relevant studies as possible, five scientific databases (ERIC, ScienceDirect, JSTOR, ProQuest, and Scopus) and one search engine (Google Scholar) were used to find papers. This search was performed using the English key words "Artificial Intelligence" AND "teaching EFL," "Artificial Intelligence" AND "EFL," "Artificial Intelligence" AND "language learning," and "ChatGPT" AND "language teaching."

Table 1. Inclusion and exclusion criteria.

Criterion	Inclusion	Exclusion
Publication date	2019–2023	Prior to 2019
Publication type	Scholarly articles from peer-reviewed journals	Book chapters, dissertations, conference proceedings, observational studies, editorials, systematic reviews, and meta-analyses
Article focus	Using AI technology in the EFL context	Any other technological context related to AI
Research method and results	Experimental studies designed to yield practical results for EFL	Reviews, opinions, discussions that did not contain empirical data analysis
Language	English	Any other language

In the initial stage of the literature search, 284 articles were retrieved using these keywords and phrases. The potential relevance of these articles was evaluated based on the inclusion criteria, and 255 articles were removed, which reduced the total number of studies to 29. Following this, the remaining 29 papers underwent further examination to identify those directly related to the incorporation of AI into EFL teaching (see Figure 1). Based on this examination, 13 articles were assessed using a quality assessment checklist and chosen for inclusion in this SLR.

**Figure 1.** Search process flowchart.

Step 3: Assessing Study Quality

In a review, it is crucial to ensure the reliability and credibility of the selected studies and evaluate their quality in addition to applying inclusion and exclusion criteria. Employing a quality assessment checklist enables researchers to establish more specific criteria for inclusion/exclusion and investigate whether discrepancies in study findings can be attributed to differences in quality (Kitchenham & Charters, 2007). It also enables the researcher to give appropriate weight to each study separately by synthesizing its results and acknowledging its importance and strength.

In this study, Kitchenham and Charters (2007) quality assessment checklist was used to evaluate the quality of the 13 studies included. Their checklist assembles a list of questions and suggests that researchers should not use all of these proposed questions but rather select the quality assessment questions that are most relevant and appropriate to their particular study context and questions. Therefore, five questions were selected from Kitchenham and Charters' checklist to evaluate the quality of the studies included in this review (as shown in Table 2).

The assessment process took place as follows: While reading the full text of each article, each of the five questions was answered, and scores were assigned to evaluate these answers on a scale ranging from 0 to 1. These answers reflect the extent to which the article met the reviewed criteria. For example, if the study clearly stated its objective, it would receive a full score of 1 for criterion Q1. Conversely, if the study did not mention its intent, it would receive a score of 0. A score of 0.5 would be given if the objective was vaguely stated. Studies with a total score below 4 out of 5 were not considered for inclusion. Table 3 presents the quality level of the 13 articles included in this review. After evaluating the articles based on the five questions, it was found that ten articles received a score of 5, one article got a score of 4, and two articles got a score of 4.5. Therefore, all 13 articles are eligible to be included in this systematic literature review.

Table 2. Quality assessment questions.

No.	Question	Scores
Q1	Are the aims clearly stated?	Yes (1) / No (0) / Partly (0.5)
Q2	Are the data collection methods adequately described?	Yes (1) / No (0) / Partly (0.5)
Q3	Is the population being studied clearly defined?	Yes (1) / No (0) / Partly (0.5)
Q4	Are the results interpreted and reported clearly?	Yes (1) / No (0) / Partly (0.5)
Q5	Do the findings highlight the impacts of incorporating AI in an EFL context?	Yes (1) / No (0) / Partly (0.5)

Source: Kitchenham and Charters (2007).

Table 3. Study quality assessment scores.

Article	Q1	Q2	Q3	Q4	Q5	Total (Level of quality)
Al-Garaady and Mahyoob (2023)	1	1	1	1	1	5
Rad, Alipour, and Jafarpour (2023)	1	1	1	1	1	5
Han et al. (2023)	1	1	1	1	1	5
Çakmak (2022)	1	1	1	1	1	5
Alsadoon (2021)	0.5	1	1	1	1	4.5
Lin and Mubarak (2021)	1	1	1	0.5	1	4.5
Xiao and Park (2021)	1	1	1	1	1	5
Dizon and Gayed (2021)	1	1	1	1	1	5
Kim et al. (2021)	1	1	1	1	1	5
El Shazly (2021)	1	1	1	1	1	5
Kholis (2021)	1	0.5	0.5	1	1	4
Qinghua and Satar (2020)	1	1	1	1	1	5
Dizon (2020)	1	1	1	1	1	5

Step 4: Summarizing the Evidence

Background details pertaining to each study reviewed are summarized in Table 4. The table includes the objective of each study, the English skill under investigation, the participants, and the AI tool used. Additionally, the data collection method and key findings are briefly summarized. The articles are organized chronologically, with the most recent study (from 2023) listed first, followed by older studies.

Step 5: Discussion (Interpreting the Findings)

Having reviewed the recent academic literature on the use of AI techniques and tools in the context of teaching EFL, we can now draw together our findings and consider the potential pedagogical implications. This discussion addresses each of the research questions in turn, providing answers based on the evidence contained in the relevant literature. To keep the scope manageable, the focus remains on EFL classroom teaching and learning in further and higher education, excluding settings with younger children and applied or specialized areas, such as professional or technical fields. The conclusion follows below, which provides some practical guidelines on how AI can be successfully integrated into this kind of EFL provision.

Research Question 1. Research question 1 sought to identify the various methods and approaches used in implementing AI technology to teach EFL and explore how these methods contribute to enhancing student learning outcomes. Student use of a chatbot appears to improve conversational practice (Alsadoon, 2021; Çakmak, 2022; El Shazly, 2021; Kim et al., 2021; Lin & Mubarak, 2021). Improvements have been observed even when students use the applications outside the classroom (Çakmak, 2022) and enhancements with mind-mapping capabilities can be beneficial (Lin & Mubarak, 2021). Negotiation for meaning can also be enhanced by pedagogical and conversational chatbot interventions (Qinghua & Satar, 2020). Chatbots can be used to prepare for roleplay work in the classroom (El Shazly, 2021) enhance vocabulary acquisition (Alsadoon, 2021) and facilitate essay revision (Han et al., 2023). These studies show that this type of chatbot technology can, and arguably should, be integrated with the standard curriculum in ways that enhance rather than replace teacher–student interaction. There is no single approach to this integration but rather many different options that can be adapted to local circumstances and targeted at a range of different language skills.

The use of a personal assistant such as Alexa can also enhance speaking skills, but, perhaps surprisingly, it did not improve listening skills (Dizon, 2020). This result may be partly due to the design of the relevant experiment, which involved only 2 small groups of 13 and 15 students and may not have been able to measure listening skills adequately. Some research suggests that the benefits of AI interventions vary according to the language level of the student (Kim et al., 2021; Qinghua & Satar, 2020). These studies offer some insight into the importance of contextual factors in the use of AI, and they suggest that needs analysis and monitoring are essential for the best results.

The use of an intelligent writing assistant has positive effects on student performance, as evidenced by fewer errors and greater lexical variation (Dizon & Gayed, 2021) as well as improved engagement and feedback literacy (Rad et al., 2023). Some positive results have been reported regarding the use of ChatGPT by instructors to identify and analyze writing errors (Al-Garaady & Mahyoob, 2023) but these benefits may be limited to the more superficial aspects of language learning. Similarly, automatic speech recognition can be used to enhance error diagnosis in pronunciation (Kholis, 2021; Xiao & Park, 2021) and this may, in turn, enhance pronunciation itself (Kholis, 2021). More research is needed to explore how AI can best be implemented for written and spoken error analysis and pronunciation in EFL.

Table 4. Overview of the studies reviewed.

No.	Authors and year	Objective of the study	Language skill	Participants	AI tool used	Data collection method	Findings
1	Al-Garaady and Mahyoob (2023)	To evaluate the effectiveness of chat generative pre-trained transformer (ChatGPT), an AI-based tool, in detecting writing errors made by EFL learners compared to human instructors	Writing	University students (Male = 54, female = 34)	ChatGPT	<ul style="list-style-type: none"> • A corpus-based research design • Human instructors analyzed texts written by students • The same written texts were analyzed using ChatGPT 	<ol style="list-style-type: none"> 1. ChatGPT primarily identified surface-level errors and struggled with identifying errors related to deep structure and pragmatics 2. Human instructors were more adept than ChatGPT at identifying complex issues in writing 3. Incorporating ChatGPT into language-learning environments enhances error analysis and improves writing skills and language proficiency
2	Rad et al. (2023)	To investigate the role of AI implementation in developing writing feedback literacy, enhancing writing engagement, and improving the overall writing outcomes of English learners	Writing	46 EFL students from an Iranian-English language institute	Wordtune (An AI-based writing app)	<ul style="list-style-type: none"> • Pre- and post-tests (Writing tasks) • Intervention: • The control group was exposed to a traditional lecture-based teaching strategy; the experimental group used an AI application (Wordtune) to practice writing skills • Writing feedback literacy scale • Writing engagement scale • Semi-structured interviews 	<ol style="list-style-type: none"> 1. The experimental group, which used the AI-based application Wordtune, demonstrated significant improvements in writing outcomes, engagement, and feedback literacy 2. Students expressed positive feedback regarding the effectiveness of Wordtune in enhancing their writing
3	Han et al. (2023)	To investigate students' perception and usage of ChatGPT in English writing through a novel learning platform called RECIPE (Revising an essay with ChatGPT on an interactive platform for EFL learners)	Writing	213 college students and 7 instructors in South Korea	ChatGPT	<ul style="list-style-type: none"> • Preliminary questionnaire (To investigate students' attitudes toward and expectations for ChatGPT) • Educational platform (RECIPE) • Interviews (To explore the platform's potential for integrating AI into the field of EFL education) 	<ol style="list-style-type: none"> 1. Students reported a positive user experience using ChatGPT 2. Students actively engaged with RECIPE, suggesting that it has the potential to enhance the learning experience
4	Çakmak (2022)	To examine the impact of chatbot-human interaction, specifically with the Replika chatbot, on students' second language (L2) speaking performance and speaking anxiety	Speaking	90 English EFL students enrolled at a state university in Turkey	Replika application	<ul style="list-style-type: none"> • Pre- and post-tests (To measure students' performance in speaking tasks and speaking anxiety) • Intervention: Conversational tasks using the chatbot Replika • - An open-ended questionnaire (To gather students' opinions on using chatbots for L2 speaking practice) 	<ol style="list-style-type: none"> 1. The chatbot had a positive impact on students' speaking performance, leading to significant improvement 2. Although students performed well with the chatbot, their speaking anxiety increased after interacting with the chatbot 3. Students reported a negative perception of the chatbot as an English conversation partner
5	Alsadoon (2021)	To investigate the influence of an interactive chatbot enhanced with vocabulary-learning tools on EFL vocabulary learning	Vocabulary	20 EFL students (aged 18-33 years) at the British Council in	AI chatbot	<ul style="list-style-type: none"> • A chatbot with four vocabulary tools: A dictionary, images, a translation tool, and a concordance • Screen recordings (Recording students' 	<ol style="list-style-type: none"> 1. The vocabulary-learning chatbot positively impacted learners' vocabulary knowledge, with the dictionary tool being the most favored and effective 2. Participants reportedly enjoyed interacting with

No.	Authors and year	Objective of the study	Language skill	Participants	AI tool used	Data collection method	Findings
		among Saudi students		Saudi Arabia		look-up behavior) <ul style="list-style-type: none"> • Pre- and post-questionnaire (To explore students' attitudes toward the chatbot) • Pre- and post-tests (To evaluate vocabulary knowledge and learning outcomes from using the chatbot) • Delayed post-tests (To check students' retention) 	the chatbot, perceiving it as a valuable resource for both authentic conversational practice and vocabulary acquisition
6	Lin and Mubarak (2021)	To investigate the effect of an integrated mind-map guided and AI chatbot approach (MM-AI) in facilitating students' speaking performance and interactions during the learning process	Speaking/ Interaction	50 EFL students at a Taiwanese university	AI chatbot	<ul style="list-style-type: none"> • Pre- and post-tests of speech • Intervention: Activities using the AI chatbot • Experimental group (Students who used MM-AI) • Control group (Students who used the conventional AI chatbot) 	1. The MM-AI approach had a positive impact on students' English-speaking performance 2. The MM-AI approach facilitated and organized interaction between robots and humans
7	Xiao and Park (2021)	To examine the effectiveness of automatic speech recognition (ASR) technology in identifying English pronunciation errors and investigate the attitudes of teachers and learners towards the use of ASR technology as both a pronunciation assessment and learning tool	Pronunciation	Five Chinese EFL learners (Ages 19-20)	ASR tool	<ul style="list-style-type: none"> • Read-aloud tests • A human-assessed test • An ASR-assessed test • Interviews (To investigate students' attitudes) 	1. The pronunciation errors identified by the ASR overlapped with those detected by human raters 2. The ASR technology successfully met the diverse pronunciation-learning needs of the learners
8	Dizon and Gayed (2021)	To assess the effects of an intelligent writing assistant tool (Grammarly) on the quality of English writing produced by Japanese students	Writing	31 university EFL students at a private Japanese university	Grammarly (An intelligent writing tool)	<ul style="list-style-type: none"> • Freewriting tasks on students' smartphones under two conditions: • Control: Writing without any aid • Experimental: Writing with the assistance of Grammarly 	1. Grammarly had a significant, positive effect on the grammatical accuracy and lexical richness of L2 students' language and helped them write more accurately
9	Kim et al. (2021)	To examine the effects of using AI chatbots in class activities according to students' proficiency levels and how they motivate and shape students' speaking experiences in the EFL classroom	Speaking	49 university students who enrolled in a general English course	AI chatbots (Replika, Andy, and Google assistant)	<ul style="list-style-type: none"> • Pre- and post-test • design (To compare students' improvement in English speaking within and between two proficiency levels) • Speaking practice with AI chatbots on students' smartphones 	1. Students showed significant improvement in the speaking tasks, and most of them spoke more proficiently after practicing with AI chatbot 2. The students improved their pronunciation, intonation, and stress
10	El Shazly (2021)	To investigate the use of AI in managing foreign language anxiety (FLA) and improving	Speaking	48 undergraduate participants in Egypt (Aged 18–	Online chatbots (Audrey, Charles, Cristal,	<ul style="list-style-type: none"> • Self-report questionnaire (To evaluate FLA) 	1. Learners experienced FLA both before and after the intervention, with slight intensification of anxiety during interactions with AI chatbots

No.	Authors and year	Objective of the study	Language skill	Participants	AI tool used	Data collection method	Findings
		foreign language speaking proficiency in an EFL class		20 years)	and Mike)	<ul style="list-style-type: none"> • Pre- and post-tests of speech (To evaluate the oral proficiency of the learners) • Intervention (The students were introduced to different AI-driven applications with web chatbots as written and oral communicative virtual partners) 	2. However, there was a significant improvement in speaking proficiency scores following the intervention, highlighting the potential of AI chatbots for enhanced communication in EFL contexts
11	Kholis (2021)	To investigate the effect of implementing English language speech assistant. ELSA speak application on students' pronunciation skills	Pronunciation	18 higher education students at Nahdlatul Ulama University of Yogyakarta (UNU), Indonesia	ELSA speak	<ul style="list-style-type: none"> • Pre- and post-tests of pronunciation • Intervention (Using ELSA speak when teaching pronunciation) • Interviews (To investigate students' feelings about using ELSA speak) • A questionnaire related to the use of ELSA Speak in teaching pronunciation 	<ul style="list-style-type: none"> • The ELSA speak application helped the students pronounce more easily and precisely • ELSA speak can support and improve students' pronunciation
12	Qinghua and Satar (2020)	To investigate chatbots' potential in foreign language educating by exploring the frequency and patterns of negotiation for eamning (NfM) in computer-mediated communication	Conversation (NfM)	Eight undergraduate and postgraduate EFL students from China	AI chatbots (Mike and Mitsuku)	<ul style="list-style-type: none"> • Interact with the chatbots for 30 minutes each, resulting in 16 chat scripts 	<ol style="list-style-type: none"> 1. Interaction with chatbots can provide learners with opportunities for NfM and thus language learning 2. Learners with lower language proficiency benefited the most from interacting with the chatbots 3. Learners with higher language proficiency expressed dissatisfaction with the chatbots and showed limited engagement in their interactions with the chatbots
13	Dizon (2020)	To investigate the use of an intelligent personal assistant (Alexa) in the classroom setting to enhance listening comprehension and speaking proficiency among EFL students	Listening and speaking	28 EFL students at a Japanese university	(Alexa)	<ul style="list-style-type: none"> • Two listening comprehension tests • A speaking proficiency test • Intervention (A 10-week in-class treatment consisting of 12 minutes of Alexa interaction each week) • Survey (To assess students' perceptions of Alexa) 	<ol style="list-style-type: none"> 1. Students demonstrated notable advancements in L2 speaking proficiency through the use of Alexa 2. There were no significant impacts of Alexa on students' listening comprehension 3. Students perceived Alexa as a useful tool for learning English.

These improvements to core skills, such as speaking, listening, and writing, were identified through quantitative analyses of student performance. Thus, they can be relied upon as a general endorsement of the integration of chatbots, automated writing evaluation, and writing assistance technologies into EFL. However, other research methods, such as attitude questionnaires, have provided quantitative results that are much more mixed. Some researchers have reported positive student attitudes toward chatbot applications (Alsadoon, 2021) while others have reported increased anxiety (Alsadoon, 2021; Çakmak, 2022).

Research Question 2. We must conclude from these decidedly mixed findings that the pedagogical implications of integrating AI techniques into EFL are bound to be varied, complex, and difficult to quantify. Results appear to vary in different settings and among individual students and teachers as well as across different levels of language ability, leading to contradictory recommendations in the relevant literature. This is, perhaps, to be expected in a field that is so new.

Overall, this analysis provides strong evidence that a range of different chatbot applications, used in a variety of different ways, can have a positive impact on pronunciation and speaking performance. This comes, however, with some caveats and limitations, such as decreased motivation in some students, increased anxiety, and benefits in one skill (e.g., speaking) but not another (e.g., listening). The benefits are likely to be more obvious in the more superficial areas of language learning, and as students become more proficient, they may be more easily bored with AI, so that the benefits are less apparent. The evidence for writing assistants of various kinds also suggests great potential and fewer negative implications, but the number of studies thus far is small, and cohort sizes tend to also be small. Therefore, the results may not yet be reliable.

4. Conclusion

This decisively mixed picture makes it difficult to provide uniformly applicable guidelines for effectively integrating AI into language classrooms. Nevertheless, the following specific guidelines point the way toward the successful integration of AI:

- It is best to aim for more basic and superficial levels and skills at first to avoid vague or unrealistic expectations.
- EFL professionals must clearly define and convey the precise goals that they aim to achieve with the use of AI (e.g., routine practice of listening, speaking, reading, or writing; vocabulary building; error analysis; and essay revision).
- Chatbots and writing assistants are a good starting point for the use of AI, but time must be devoted to familiarizing instructors and students with the benefits of their use as well as any potential limitations or disadvantages.
- Attention must be paid to integrating AI techniques with other teaching and learning methods and equipment inside and outside the classroom.
- More empirical studies are needed in order to test the theories and results presented in the slowly emerging research in this exciting field.

References

- Al-Garaady, J., & Mahyoub, M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journals, Special Issue on Call*, 9, 1-15. <https://doi.org/10.24093/awej/call9.1>
- Alsadoon, R. (2021). Chatting with AI Bot: Vocabulary learning assistant for Saudi EFL learners. *English Language Teaching*, 14(6), 135-157.
- Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52-62. <https://doi.org/10.61969/jai.1337500>
- Bonfield, C. A., Salter, M., Longmuir, A., Benson, M., & Adachi, C. (2020). Transformation or evolution? Education 4.0, teaching and learning in the digital age. *Higher Education Pedagogies*, 5(1), 223-246. <https://doi.org/10.1080/23752696.2020.1816847>
- Çakmak, F. (2022). Chatbot-human interaction and its effects on EFL students' L2 speaking performance and anxiety. *Novitas-ROYAL Research on Youth and Language*, 16(2), 113-131.
- Colquhoun, H. L., Levac, D., O'Brien, K. K., Straus, S., Tricco, A. C., & Perrier, L. (2014). Scoping reviews: Time for clarity in definition, methods, and reporting. *Journal of Clinical Epidemiology*, 67(12), 1291-1294. <https://doi.org/10.1016/j.jclinepi.2014.03.013>
- Cordeschi, R. (2007). AI turns fifty: Revisiting its origins. *Applied Artificial Intelligence*, 21(4-5), 259-279. <https://doi.org/10.1080/08839510701252304>
- Cowie, N., & Alizadeh, M. (2022). The affordances and challenges of virtual reality for language teaching. *International Journal of TESOL Studies*, 4(3), 50-56. <https://doi.org/10.46451/ijts.2022.03.05>
- Cullinan, J., Flannery, D., Harold, J., Lyons, S., & Palcic, D. (2021). The disconnected: COVID-19 and disparities in access to quality broadband for higher education students. *International Journal of Educational Technology in Higher Education*, 18(1), 26. <https://doi.org/10.1186/s41239-021-00262-1>
- Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., & Braasch, J. (2022). Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation. *Computer Assisted Language Learning*, 35(9), 2332-2360. <https://doi.org/10.1080/09588221.2021.1879162>
- Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology*, 24(1), 16-26.
- Dizon, G., & Gayed, J. M. (2021). Examining the impact of grammarly on the quality of mobile L2 writing. *Jalt Call Journal*, 17(2), 74-92. <https://doi.org/10.29140/jaltcall.v17n2.336>
- El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, 38(3), e12667. <https://doi.org/10.1111/exsy.12667>
- Godwin-Jones, R. (2017). Scaling up and zooming in: Big data and personalization in language learning. *Language Learning and Technology*, 21(1), 4-15.
- Han, J., Yoo, H., Kim, Y., Myung, J., Kim, M., Lim, H., & Oh, A. (2023). *RECIPE: How to integrate ChatGPT into EFL writing education*. Paper presented at the In Proceedings of the Tenth ACM Conference on Learning@ Scale.
- Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, 96(3), 118-121. <https://doi.org/10.1258/jrsm.96.3.118>
- Kholis, A. (2021). Elsa speak app: Automatic speech recognition (ASR) for supplementing English pronunciation skills. *Pedagogy: Journal of English Language Teaching*, 9(1), 01-14. <https://doi.org/10.32332/joelt.v9i1.2723>
- Kim, H.-S., Cha, Y., & Kim, N. Y. (2021). Effects of AI chatbots on EFL students' communication skills. *Korean Journal of English Language and Linguistics*, 21, 712-734.
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering version 2.3. *Engineering*, 45(4ve), 1051.

- Lin, C.-J., & Mubarak, H. (2021). Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society*, 24(4), 16-35.
- Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology*, 50(6), 2824-2838. <https://doi.org/10.1111/bjet.12861>
- Qinghua, Y., & Satar, M. (2020). English as a foreign language learner interactions with chatbots: Negotiation for meaning. *International Online Journal of Education and Teaching*, 7(2), 390-410.
- Rad, H. S., Alipour, R., & Jafarpour, A. (2023). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of wordtune application. *Interactive Learning Environments*, 1-21. <https://doi.org/10.1080/10494820.2023.2208170>
- Roschelle, J., Lester, J., & Fusco, J. (2020). *AI and the future of learning: Expert panel report [Report] digital promise*. Retrieved from <https://eric.ed.gov/?id=ED614308>
- Schmidt, T., & Strasser, T. (2022). Artificial intelligence in foreign language learning and teaching: A CALL for intelligent practice. *Anglistik: International Journal of English Studies*, 33(1), 165-184. <https://doi.org/10.33675/angl/2022/1/14>
- Singh, S. V., & Hiran, K. K. (2022). The impact of AI on teaching and learning in higher education technology. *Journal of Higher Education Theory & Practice*, 12(13), 135-148. <https://doi.org/10.33423/jhetp.v22i13.5514>
- Taguchi, N. (2015). "Contextually" speaking: A survey of pragmatic learning abroad, in class, and online. *System*, 48, 3-20. <https://doi.org/10.1016/j.system.2014.09.001>
- Tawfik, G. M., Dila, K. A. S., Mohamed, M. Y. F., Tam, D. N. H., Kien, N. D., Ahmed, A. M., & Huy, N. T. (2019). A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Tropical Medicine and Health*, 47, 1-9. <https://doi.org/10.1186/s41182-019-0165-6>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., . . . Weeks, L. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467-473.
- Xiao, W., & Park, M. (2021). Using automatic speech recognition to facilitate English pronunciation assessment and learning in an EFL context: Pronunciation error diagnosis and pedagogical implications. *International Journal of Computer-Assisted Language Learning and Teaching*, 11(3), 74-91. <https://doi.org/10.4018/ijcallt.2021070105>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>