

A systematic review of conversational AI tools in ELT: Publication trends, tools, research methods, learning outcomes, and antecedents

Wan Yee Winsy Lai, Ju Seong Lee*

The Education University of Hong Kong, Hong Kong



ARTICLE INFO

Keywords:

Conversational AI
AI chatbot
AI-assisted language learning
ELT
English as a foreign language
Ecological system theory
Systematic review

ABSTRACT

This review analyzed the trends in conversational AI tools in ELT from January 2013 to November 2023. The study examined 32 papers, focusing on publication trends, tool types, research methods, learning outcomes, and factors influencing their use. Findings revealed a gradual increase in publications, with 4 (12%) from 2013 to 2021, 13 (41%) in 2022, and 15 (47%) in 2023. All studies (100%) were conducted in Asian EFL contexts. Among the AI chatbots, *Google Assistant* (25%) was the most widely used. Quasi-experimental (45%) and cross-section (41%) research designs were commonly employed. Mixed-method (50%) approaches were prevalent for data collection and analysis. Conversational AI yielded positive outcomes in affective (43%) and cognitive skills (41%). The main factors influencing user perceptions or behaviors were individual (47%) and microsystem layers (31%). Future studies should (a) include diverse contexts beyond Asia, (b) consider the use of up-to-date tools (e.g., *ChatGPT*), (c) employ rigorous experimental designs, (d) explore behavioral learning outcomes, and (e) investigate broader environmental factors. The systematic review enhances current knowledge of recent research trends, identifies environmental factors influencing conversational AI tools concentrating in ELT, and provides insights for future research and practice in this rapidly evolving field.

1. Introduction

Conversational agents, computer programs capable of understanding and responding to human language, have found applications in various domains, such as healthcare, e-commerce, and education (Kulkarni et al., 2019). Artificial Intelligence (AI) chatbots, powered by technologies like natural language processing, engage learners in real-time interactions, leading to significant improvements in language education (Kohnke et al., 2023; Ou et al., 2024). Many studies have shown that students benefit from interacting with AI chatbots in language learning, improving their pronunciation (Liakin et al., 2015), grammar (Al-Kaisi et al., 2019), and vocabulary (Chen et al., 2020). There have been several reviews discussing the roles and advantages of AI chatbots, including their role as conversational partners (Ji et al., 2023), their ability to create an affective and open communication environment (Huang et al., 2022), and their incorporation of empathy to elicit appropriate emotional responses for students (Zhai & Wibowo, 2022).

In the field of English language teaching (ELT), conversational AI tools have been increasingly integrated to provide an engaging environment for students. These tools are individualized, interactive, and

immersive, enhancing students' confidence in speaking English (Ruan et al., 2021), sustaining their interest in reading (Liu et al., 2022), and improving their writing abilities through personalized feedback (Hwang et al., 2023). While some researchers reviewed AI tools in ELT and presented linguistic outcomes in vocabulary, reading, writing, and communication skills (Klimova et al., 2023; Liang et al., 2023; Sharadgah & Sa'di, 2022; Zhai & Wibowo, 2023), there is a lack of reviews specifically focusing on conversational AI tools in the ELT context.

Therefore, three areas need improvement in the existing review of conversational AI tools in ELT. First, there is a need for more up-to-date information on publications and the development of conversational AI tools. For instance, six systematic review papers on AI applications (e.g., AI chatbots and other applications) have been published (Huang et al., 2022; Ji et al., 2023; Klimova et al., 2023; Liang et al., 2023; Sharadgah & Sa'di, 2022; Zhai & Wibowo, 2023), the latest search of articles was conducted up till August 2022 (Zhai & Wibowo, 2023), resulting in a lack of recent developments, such as the launch of *ChatGPT* in November 2022. Second, there is a scarcity of systematic reviews specifically focusing on conversational AI tools in ELT. To date, most related reviews have either focused on AI applications in the ELT (Liang et al., 2023;

* Corresponding author. Department of English Language Education, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong.
E-mail address: jslee@edu.hk (J.S. Lee).

Sharadgah & Sa'di, 2022) or conversational AI applications in language learning (Huang et al., 2022; Ji et al., 2023), leaving a significant gap regarding the scope of conversational AI tools in ELT. Lastly, the understanding of conversational AI tools has been limited to individual characteristics, neglecting the broader environmental contexts that influence students' perception and usage of such tools. Existing reviews on conversational AI applications have mainly adopted a student-centric approach, emphasizing students' background and language outcomes while disregarding more macro-level environments (Huang et al., 2022; Zhai & Wibowo, 2023).

To address these gaps, this review aims to provide a comprehensive and up-to-date analysis of conversational AI tools in ELT from January 2013 to November 2023. It will cover (a) publication trends (b) AI tool trends, (c) research design types (e.g., experimental or cross-sectional), (d) data collection and analysis strategies (e.g., quantitative or qualitative), (e) language learning outcomes, and (f) influential factors ranging from individualistic factors to macro-environment factors.

2. Literature review

2.1. Conversational AI tools

In the development of AI technology, the initial phase relied on machine learning (ML) algorithms to predict outcomes based on past data and behaviors (Law, 2024). As ML evolved, Generative AI (GenAI) emerged, leveraging natural language processing (NLP) to facilitate machines to understand, interpret, and generate human-like language (Roumeliotis & Tselikas, 2023). Within NLP, large language models (LLMs) are particularly noteworthy. These models are trained on vast and varied datasets, allowing them to perform complex language tasks. Notable examples include GPT-3 and GPT-4, which power ChatGPT.

Conversational AI tools refer to chatbots powered by AI that engage in dialogue with users using LLMs, NLP, and ML algorithms. These tools analyze user input and generate text or speech responses in a human-like and conversational manner, mimicking a real conversation (Ou et al., 2024; Smutny & Schreiberová, 2020). Conversational AI tools can be divided into text-based and voice-based chatbots. Text-based chatbots communicate through written messages, storing users' input and generating AI-driven responses that resemble human conversation. Examples include models like the ChatGPT-4 model and Google Bard (Yang et al., 2022). By contrast, voice chatbots interact with users through spoken commands and responses, often found in smartphone applications, such as Siri, Amazon Alexa, and Google Assistant (Alnafaei et al., 2021). For the purpose of this review, conversational AI tools are defined as chatbots capable of recognizing and generating text or voice responses, simulating human-like interactions, and are accessible to users through mobile and web platforms.

2.2. Previous studies on conversational AI applications and ELT

The increasing popularity and accessibility of AI tools have led to a rise in students using them for learning purposes (Alharbi, 2023). When AI chatbots are introduced in ELT classrooms, students have positive experiences, finding them easy to use and comfortable (Yang et al., 2022). In self-directed English learning, students have a positive attitude toward ChatGPT (Liu & Ma, 2023). EFL teachers appreciate conversational AI tools for their ability to provide personalized teaching materials, real-time feedback, and human-like responses, which save time, motivate students, and enhance language acquisition (Jia et al., 2022; Mohamed, 2023; Qasem et al., 2023). However, some teachers express concerns about students becoming overly dependent on these tools, which may result in reduced human interaction (Qadir, 2023), diminished critical thinking skills (Mohamed, 2023), potential cheating (Mhlanga, 2023), and uncertainties about the accuracy of chatbot responses (Kohnke et al., 2023).

Empirical studies have shown improvements in various aspects of

ELT through the use of conversational AI tools. These improvements include enhanced reading abilities (Liu et al., 2022), improved writing skills (Lin & Chang, 2020), better listening comprehension (Chien et al., 2022), enhanced speaking proficiency (Lin & Mubarok, 2021), increased vocabulary acquisition (Ruan et al., 2021), and improved grammar knowledge (Nghi et al., 2019). AI chatbots offer a wide range of vocabulary and expressions, thereby expanding students' exposure to the English language (Huang et al., 2022). Regular practice in conversational skills with immediate feedback contributes to the development of language proficiency (Huang et al., 2022; Kim et al., 2019; Yang et al., 2022). Additionally, AI chatbots can adjust the difficulty level of tasks and offer personalized learning materials to cater to individual learner needs (Kuhail et al., 2023). As AI continues to advance, there are new opportunities for integrating chatbots into classrooms (Belda-Medina & Calvo-Ferrer, 2022).

While studies have shown significant improvements in students' English proficiency, a few have highlighted potential negative impacts and challenges of using AI chatbots in ELT. First, despite the urgent issue of ethical use, only a limited number of studies have addressed it. As AI becomes more popular in ELT, it is crucial for educators to promptly identify and manage the ethical and appropriate use of AI chatbots in their institutions (Escalante et al., 2023). In second language teaching, it is essential to train educators in ethical principles to guide students in tasks designed for AI chatbots, like ChatGPT (Kohnke et al., 2023). Similarly, Jiao et al. (2023) found that by regulating teaching pedagogy, students could become aware of AI ethics as both users and researchers.

Second, some users, such as parents, might feel threatened by the intelligence of AI bots and prefer applications with lower autonomy (Lin et al., 2021). However, designing AI bots with an engaging appearance tailored to increase children's engagement could reduce parents' fears and concerns about physical risks (Lin et al., 2023; Louie et al., 2021). Third, Wang, Gupta, et al. (2022) suggested that if an AI chatbot is portrayed as a teacher with a high teaching presence, this perceived high authority could be intimidating. As a result, students might find interactions with AI chatbots less friendly and more stressful, negatively impacting their EFL learning. This phenomenon has also been observed in other AI applications (Randall, 2020).

2.3. Previous review studies on AI, language learning, and ELT

Several review studies have been conducted to explore the use of AI in language education and ELT. Liang et al. (2023) conducted a systematic review analyzing 71 studies from 1989 to 2020, focusing on using AI in language education. They found that the most common outcomes were vocabulary, reading, and writing, and the most popular technologies used were the Intelligent Tutoring System (ITS) and Natural Language Processing. However, their review solely focused on language learning outcomes and did not consider other affordances of AI. To address this limitation, Huang et al. (2022) reviewed 25 studies till 2020 that examined the technological, pedagogical, and social affordances of chatbots and language learning. They highlighted the roles of chatbots as timely, easy-to-use language learning tools, personalized interlocutors, simulators, and advisors. However, their review included chatbots that were not AI-supported, and neither review specifically concentrated on ELT. Sharadgah and Sa'di (2022) narrowed the focus to ELT and AI technology, reviewing 64 studies from 2015 to 2021. They explored trends related to students' English proficiency, attitudes toward AI, autonomous learning, and teachers' efficiency, primarily in higher education. However, their review encompassed a broad range of AI technologies, and similar to the previous reviews, it placed more emphasis on students rather than stakeholders involved in ELT (Huang et al., 2022; Liang et al., 2023). To address the limitations of the previous reviews, Ji et al. (2023) conducted a review that examined the collaborative roles between conversational AI tools and language teachers. They analyzed 24 empirical studies from 2015 to 2021 and shed light on the influence of

teachers in designing, evaluating, and making decisions when integrating conversational AI tools into the classroom. However, their review focused on languages in general rather than specifically on ELT. Similarly, Klimova et al. (2023) conducted a more recent review that investigated the implications of emerging technologies, including AI applications, on foreign language learning from both students' and teachers' perspectives. They analyzed 14 studies from 2018 to 2022 and highlighted a lack of chatbots designed for foreign language learning.

To address the gap in reviews on ELT and conversational AI applications, Zhai and Wibowo (2023) investigated 28 studies from January 2013 to August 2022. Their review explored the usage of AI chatbots in enhancing EFL university students' interactional competence. They emphasized the ELT context and identified six key dimensions that influence the application of AI dialogue systems: technological integration, task design, learners' engagement, learning objectives, novelty effect, and system limitations. However, their review may not fully encompass the latest applications, as it was conducted before the emergence of popular tools like ChatGPT in the EFL context (Bin-Hady et al., 2023; Liu, 2023).

Among the studies, a majority of reviews conducted their search with the PRISMA framework in selecting the journal articles from large-scale databases, such as Web of Science, Scopus, and Google Scholar. While a few reviews included conference proceedings (Huang et al., 2022) and dissertations (Ji et al., 2023), Zhai and Wibowo (2023) explained their exclusion of editorial and conference proceedings to reduce potential bias. Furthermore, a limited selection of keywords in the search strings was observed, ranging from 4 to 10 (Huang et al., 2022; Ji et al., 2023; Klimova et al., 2023; Sharadgah & Sa'di, 2022). Zhai and Wibowo (2023) started the search with variations related to EFL, AI, interactional competence, and university education, however, zero data was retrieved, hence, they revised and simplified the search strings, which reflected the stage of AI chatbots in EFL was still in infancy stage during their search in 2022. Similarly, Law (2024) opted for a scoping review rather than a systematic review due to inadequate literature on generative AI in language education when they conducted the search in July 2023.

2.4. Purpose of the study

The objective of this study is to address several research gaps pertaining to conversational AI tools in the context of ELT. Existing literature reviews have identified three main areas where further investigation is warranted.

First, there is a notable dearth of up-to-date information regarding the latest publications and developments in conversational AI tools. The most recent search on conversational AI tools in language education was conducted in 2022, rendering it insufficient to capture subsequent advancements in the field. Particularly, the launch of ChatGPT has had a transformative impact on the ELT landscape (Kohnke et al., 2023). In under six months, an updated model, ChatGPT-4, was launched in March 2023, supported by a multimodal model to process both text and images (Shalevska, 2023). As the AI technology landscape is constantly evolving, a more up-to-date review is imperative to ensure accurate tracking of the latest tools and publications, facilitating their effective integration into ELT practices. Consequently, the current review presents an extensive review of conversational AI tools in the ELT context from January 2013 to November 2023, including (a) publication trends in years, journal titles, geographical context, and participant's background and (b) trends of which conversational AI tools were used, formulated as research questions 1 and 2.

Secondly, there exists a paucity of systematic reviews specifically focusing on conversational AI tools in the ELT context. While prior research has explored the broader domain of AI and language education, a notable gap persists in examining conversational AI tools exclusively in the ELT setting (Sharadgah & Sa'di, 2022). Given the global significance of English as a lingua franca for academic, professional, and social

purposes, understanding the impact of AI chatbots on ELT assumes paramount importance (Sharadgah & Sa'di, 2022; Yu & Nazir, 2021). Moreover, the majority of AI chatbots (e.g., Google Assistant and ChatGPT) have been developed in English-speaking countries. For instance, in the case of the GPT-3 model in ChatGPT, 93% of the training data was primarily in English, while the remaining 7% was in other languages (Brown et al., 2020); the more influential the language model, the better the performance of the bot (Bender et al., 2021), thus, conducting a review specifically in ELT context would enable learners to potentially engage in a higher quality of interactions generated from a richer database. Furthermore, conversational AI tools hold the potential to foster educational innovation in ELT by facilitating novel teaching practices, individualized learning experiences, and heightened learner engagement (Fitria, 2021; Wang, Gupta, et al., 2022). Hence, the current review narrows the scope to conversational AI tools and ELT context to present (1) research trends in design types (e.g., cross-sectional and experimental) and (2) research trends in data collection strategies used (e.g., qualitative, quantitative and mixed methods), and (3) language learning outcomes in the ELT context exclusively, designed as research questions 3, 4 and 5 respectively.

Lastly, there is a lack of understanding of the macro-environmental factors that influence the perception and usage of AI chatbots in ELT. Several factors have been identified as impacting the successful implementation of AI chatbots in ELT, including students' perceived usefulness of the chatbots (Liu & Ma, 2023), teachers' perception of their effectiveness in preparing language activities (Chocarro et al., 2023; Nguyen, 2023; Ulla et al., 2023), the complexity of tools (Du & Gao, 2022), and concerns about students' development of critical thinking skills (Mohamed, 2023). Government policies also play a role in the adoption of conversational AI tools in ELT. For instance, in the United States, the New York Department of Education has prohibited the use of ChatGPT in academic settings due to concerns about academic integrity (Mohamed, 2023). By contrast, the Korean government has collaborated with research institutions and an Internet company to develop an AI chatbot called 'AI-Pengtalk' to enhance English communication practices for learners (Um et al., 2023). To date, while some reviews have attempted to include teachers in their scope (Ji et al., 2023; Klimova et al., 2023), most existing reviews have primarily focused on the individual level, such as the student's learning outcomes (Sharadgah & Sa'di, 2022; Zhai & Wibowo, 2023), neglecting the broader environmental contexts, including parents, teachers, schools, social norms, and policies. This is important because these macro-level environmental factors can have a significant impact on the students' decisions to use AI chatbots in ELT, both directly and indirectly. Previous studies have adapted Bronfenbrenner's ecological systems theory to identify influential factors in students' emotions in online classrooms (Kruk et al., 2022) and perception of informal digital learning of English (Guo & Lee, 2023). Bronfenbrenner (1979) conceptualized influences from the environment as nested structures, with each layer contained within the next, forming an interconnected ecological environment surrounding an individual. These layers include the individual, micro-system, mesosystem, exo-system, macro-system, and chrono-system. Consequently, the current review aims to provide a comprehensive overview of the influential factors across all layers of using conversational AI tools, drawing on the lens of the ecological system theory in addressing research question 6. Fig. 1 illustrates the proposed influential factors, ranging from internal to external layers: individual (e.g., student's personality and tech-savviness), microsystem (e.g., family and teachers), mesosystem (e.g., school policies and online communities), exosystem (e.g., parents and teachers' AI proficiency), macrosystem (e.g., government language policies and social norms), and chronosystem factors (e.g., the impact of the pandemic). Specifically, the review will address the following research questions (RQ):

- RQ1. What are the publication trends regarding conversational AI tools in the field of ELT?

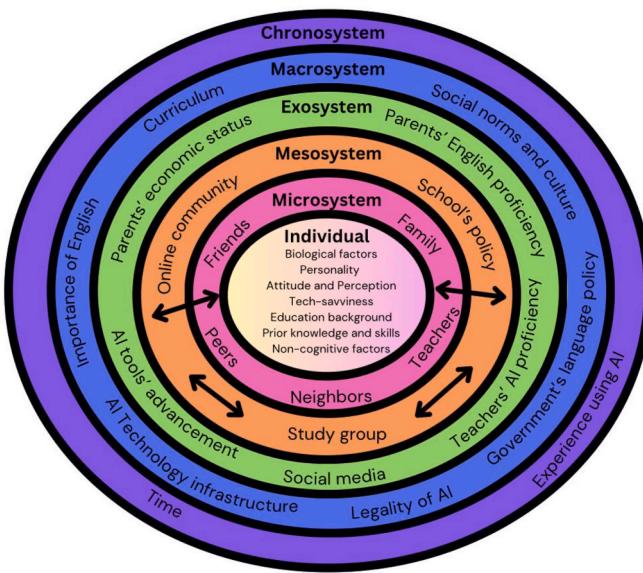


Fig. 1. Potential influential factors from Ecological system theory.

- RQ2. How are conversational AI tools being integrated into English language learning?
- RQ3. What types of research designs are commonly employed in studies focusing on conversational AI tools in ELT?
- RQ4. What data collection and analysis strategies are commonly used in research on conversational AI tools in the ELT context?
- RQ5. What are the observed learning outcomes associated with the use of conversational AI tools in ELT?
- RQ6. From an ecological perspective, what factors influence students' and teachers' perceptions and usage of conversational AI tools in the ELT context?

3. Methodology

3.1. Search strategy and literature retrieval

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). The search strategy involved three categories: (a) conversational tools, (b) AI technology, and (c) English language context. The search string included 16 interchangeable keywords related to conversational tools from previous studies and was expanded to ensure comprehensive coverage in comparison to existing reviews (Appendix A).

The search string used was as follows: ("bot" OR "chatbot" OR "conversation"*) OR "agent" OR "application" OR "partner" OR "software" OR "dialogue agent" OR "dialogue system" OR "instant messag*" OR "language tutor" OR "learning assistant" OR "personal assistant" OR "robot" OR "tutor" OR "personal" OR "agent") AND "English" AND ("Artificial Intelligence" OR "AI" OR "Intelligen*"). To maximize the number of eligible studies, four popular data sources were gathered from existing reviews and utilized: Web of Science, Scopus, ProQuest, and Google Scholar. To ensure high-quality research, only articles published in Social Science Citation Index (SSCI) and Arts & Humanities Citation Index (A&HCI) journals were included (Guo & Lee, 2023). With reference to existing reviews, the review shall exclude editorials, opinion pieces, and dissertations to reduce potential bias (Zhai & Wibowo, 2023).

3.2. Inclusion and exclusion criteria

The PRISMA flowchart (Fig. 2) outlines the inclusion and exclusion criteria. A total of 561 records from January 2013 to November 2023

were identified from four databases. After removing duplicates and records not published in SSCI and A&HCI journals 253 records remained. From the titles and abstracts, 161 studies unrelated to language learning (e.g., health science) were removed, resulting in 92 papers. To prevent omissions, articles focusing on languages other than English were included if they also involved English language learning. After a full-text review, 60 papers were excluded based on the following criteria:

1. Articles focusing on language contexts other than ELT ($n = 8$).
2. Articles without empirical data, such as reviews, commentary, and dissertations ($n = 8$).
3. Articles that did not include the implementation of conversational AI tools or chatbots that were not AI-supported ($n = 43$).
4. Articles not written in English ($n = 1$).

Ultimately, 32 studies were retained for the current review (Appendix B).

3.3. Coding framework and analysis procedures

The coding scheme was developed based on the research questions and is presented in Appendix C. The classification of the coding process is depicted below:

1. Research question 1: The frequency of studies was recorded in terms of years, journal titles, and geographical location. The unspecified group of participants' educational backgrounds included parents and educators.
2. Research question 2: The identification excluded learning concepts and AI tools that do not have conversational settings, such as automated writing evaluation tools (e.g., Grammarly), speech recognition tools (e.g., Elsa Speak), and language learning apps (e.g., Duolingo).
3. Research question 3: The research design types were labeled as cross-sectional (one-time point without intervention; Levin, 2006), quasi-experimental (interventions without randomization; Mackey & Gass, 2016), and experimental (interventions with random assignment; Mackey & Gass, 2016). Initially, a 'non-experimental' label was considered for categorization, but it was determined that this label did not apply to the studies selected for the current analysis (Thompson & Panacek, 2007).
4. Research question 4: The data collection and analysis strategies were categorized as quantitative (testing the strength among variables; Creswell, 2015), qualitative (preserving an individual's experience; Creswell & Poth, 2016), and mixed methods (combining both quantitative and qualitative methods).
5. Research question 5: The learning outcomes were classified into the affective domain ("integration of beliefs, ideas, and attitudes"; e.g., emotions; Krathwohl et al., 1964, p. 184), the behavioral domain (use of skill areas; e.g., increase in reading amount; Snowman et al., 2011), and the cognitive domain (knowledge, comprehension, application, analysis, synthesis, and evaluation; e.g., language proficiency; Krathwohl, 2002; Krathwohl et al., 1964). While some may argue that willingness to communicate is a linguistic improvement falling under the cognitive domain, Krathwohl et al. (1964) suggested that a student's willingness to respond is linked to whether they actually take action after learning they can do it. In contrast, the cognitive domain focuses on determining if the student can perform the action. Therefore, we classified willingness to communicate as an affective outcome, emphasizing the emotional and motivational aspects involved.
6. Research question 6: Influential factors were coded based on the six layers of Bronfenbrenner's ecological system theory (Bronfenbrenner, 1974, 1979). These layers include the individual layer (individual characteristics; e.g., personality), microsystem layer (immediate effect on the individual; e.g., teachers), mesosystem layer (interaction between Microsystems; e.g., schools for

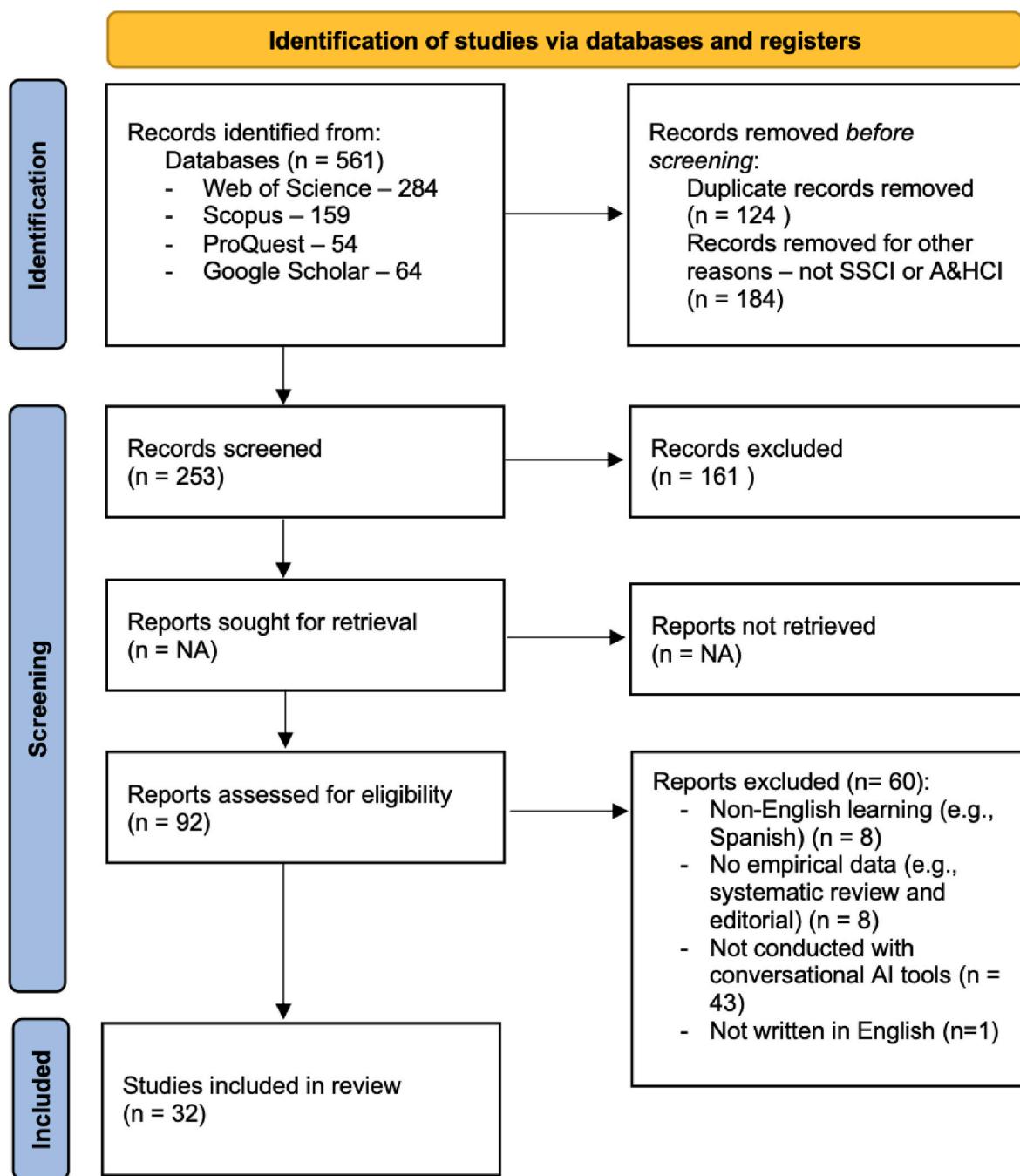


Fig. 2. PRISMA flow diagram of the study selection.

teachers' intercommunication), exo-system layer (indirect impact from social structures; e.g., teachers' digital literacy), macrosystem layer (institutional patterns of cultural 'blueprints'; e.g., national policies), and chronosystem layer (environmental changes over time; e.g., pandemic).

The coding references were derived from systematic reviews on AI tools or chatbots in education, with additional scopes such as research designs and influential factors. Fifteen studies were initially selected for preliminary coding by the first author and an external reviewer experienced in AI conversational applications, which reached inter-reliability at 85%. After comparing the preliminary coding and resolving any differences, the rest of the studies were coded accordingly.

4. Findings and discussion

The purpose of this review is to explore the trends and implications of using conversational AI in English Language Teaching (ELT) based on 32 studies conducted in the past decade. The review addresses several research questions: the latest trends in publication and conversational AI applications, research trends on design types and data collection, ELT learning outcomes, and influential factors in micro and macro environments.

4.1. Publication trends

4.1.1. Publication trends by publication years and journals

Among the 32 studies from 2013 to 2023, the chart shows a

trajectory trend in publications, from a slow start in 2016–2021, with four studies at approximately 12% to a surge of 13 studies (41%) in 2022 and reaching its pinnacle in 2023 with 15 publications at 47% (Fig. 3). This trend supports that conversational AI tools are gaining researchers' and educators' attention rapidly, in particular the last two years, in ELT communities (Mohamed, 2023). A potential reason is the rapid development and diversity of conversational AI tools which will be discussed in Section 4.2.

Fifteen journals published conversational AI and ELT studies in the past decade (Fig. 4). Computer Assisted Language Learning and Interactive Learning Environments tiered at publishing six studies, with almost 40% share, followed by Language Learning & Technology with four studies at 13% and Education and Information Technologies with three studies at 10%. Among the six studies published on Computer Assisted Language Learning, an equal split of 50% was conducted in K-12 education and universities. In comparison, four out of six articles on Interactive Learning Environments were conducted in universities and two in K-12 education. Among the selected studies, 86% of the studies were published in Q1 journals, which acknowledges the emphasis on AI-integrated ELT trends.

4.1.2. Publication trends by geographical location

Regarding participants' geographic location, almost 60% were conducted in Chinese-speaking countries and regions, with the highest number from Taiwan (35%) and, subsequently, Mainland China (15%), and Hong Kong SAR (9%) (Fig. 5). Korea had the second highest number, at 15%. Tied with Hong Kong, studies with unspecified countries were at 9%, followed by Iran at 5% and 3% in Indonesia, Turkey, Saudi Arabia, and Japan. Amidst the studies that stated the location, we unraveled a research trend that all studies were conducted in Asia, where English is a second or foreign language (EFL). With all studies conducted in the EFL context, the geographic trend highlights the importance and eagerness of researchers and educators to incorporate AI chatbots into EFL education. Their studies supported the belief that AI chatbots could support EFL students' learning, especially in oral proficiency (Bin-Hady et al., 2023; Kim et al., 2022; Tai, 2022), pronunciation (Feng & Wang, 2022) and lowering foreign language speaking anxiety (Chen, 2022).

4.1.3. Publication trends by education background

Almost half of the studies were conducted in higher education, with 33% of undergraduate and 11% of postgraduate students (Fig. 6). With almost half of the studies conducted in higher education settings, this trend can be explained by three observations. First, one of the common prerequisites is for participants to have access to computers or smartphones to utilize the AI chatbots (Ebadi & Amini, 2022), and it is more

common for university students to own a smartphone than for K-12 students. Second, although it is not a prerequisite, almost two-thirds of participants of Gonulal's study (2023) had prior experience using Intelligent Personal Assistants (IPA) for education and entertainment purposes, this group of students showed an advantage of higher usability ratings when using IPA to learn English, which demonstrated students who had access to smart devices and AI experience would present better learning outcomes. Third, convenience sampling is involved in some selected studies which will be discussed in Section 4.3. K-12 education comprised 38%, with an equal split of 19% in primary and secondary education. 17% of studies were conducted with parents and university educators, with unspecified education backgrounds, which showcased a need to focus on stakeholders apart from students.

4.2. Publication trends in AI tools

Fig. 7 presents the trend in the choice of conversational AI tools related to ELT. Google Assistant was the most popular tool in the last decade ($n = 8$), accounting for 25%, succeeded by self-developed AI bot ($n = 7$) at 22%, and ChatGPT ($n = 4$) at 13%. While six studies (19%) did not specify the chatbots, their conversational function, and AI technology were depicted in the studies to be included in the review.

There are three notable highlights. First, despite the limitation that Google Assistant and ChatGPT were not explicitly designed for educational purposes, many researchers acknowledged the necessity for experimentation with these tools in ELT. This could be explained that they perceived them as potential solutions in the EFL context. Accessibility might have contributed to their selection; Google Assistant, in particular, was available across Asia, which offered ease to introduce in ELT classrooms.

Second, we noticed the relation between the launch year of the AI tools and the publication years. Although ChatGPT did not exhibit the highest usage among selected studies, it accounted for 13% shortly in 2023 after its launch in November 2022. We foresee an upward trend in studies using ChatGPT. However, since it has not been officially launched in some countries, some researchers might prefer tools accessible in their countries, such as Google Assistant.

Third, we recognized self-developed chatbots being introduced for ELT uses. AI chatbots, such as Alexa and Google Assistant were built to promote user convenience by controlling smart home devices as IPAs instead of educational purposes. In addition, several chatbots, including Google Assistant, were created in the United States, where English is their first language. A few studies found comprehensibility issues provoked by the accents' indifference. IPAs had difficulties in understanding EFL students' utterances (Tai, 2022), especially students with lower English proficiency (Chen et al., 2023), which demotivated students to speak English (Jeon, 2022). Thus, researchers urged to develop their applications catering to EFL users (Chen et al., 2023). With the technology progression, a few researchers utilized the language models from Open AI (Hwang et al., 2023) and Google (Jeon, 2022; Kim et al., 2022; Liu et al., 2022; Yang et al., 2022) to create chatbots to accommodate EFL learners. Besides, government bodies, such as the South Korean government created AI-Pengtalk for EFL learners in primary school to enhance their proficiency (Um et al., 2023).

4.3. Research trends on design types

Fig. 8 depicts the distribution of research designs as follows: 45% quasi-experimental, 41% cross-sectional, and 14% experimental studies. The quasi-experimental method was the most popular research approach, accounting for almost 50%.

Given the constraints of class settings, researchers found it feasible to assign experimental and control groups using the existing class structure, as most interventions happened during regular class hours (Chien et al., 2022; Hwang et al., 2023; Lin & Mubarok, 2021; Liu et al., 2022). A few researchers also served as teachers which allowed them to

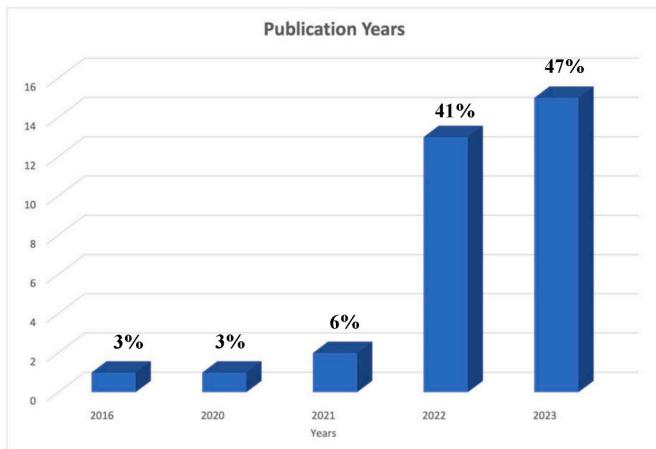


Fig. 3. Publication trend by years.

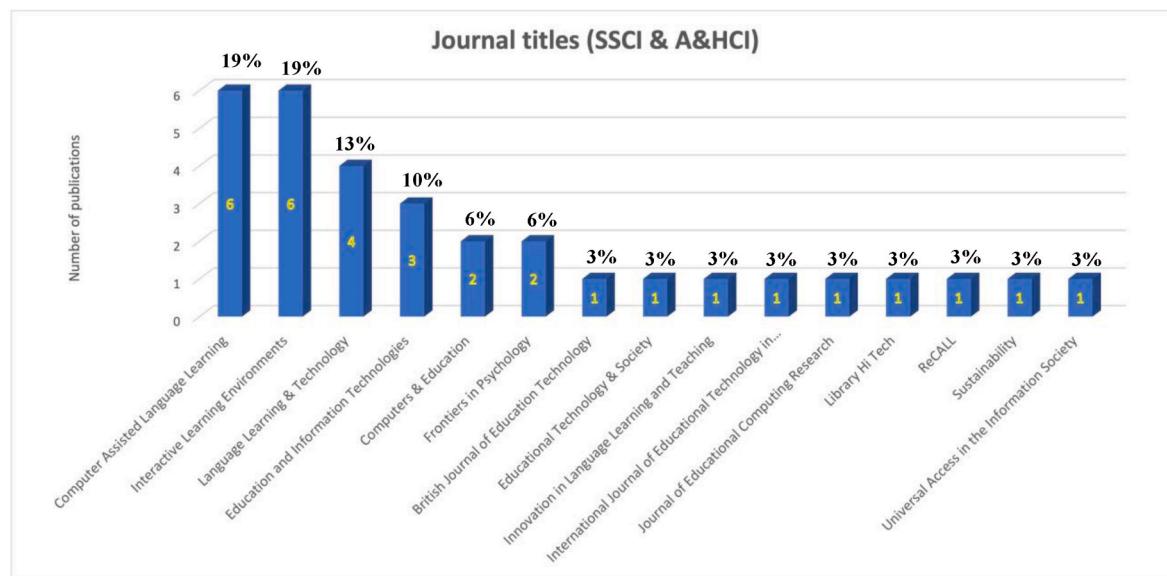


Fig. 4. Publication trend in journal titles.

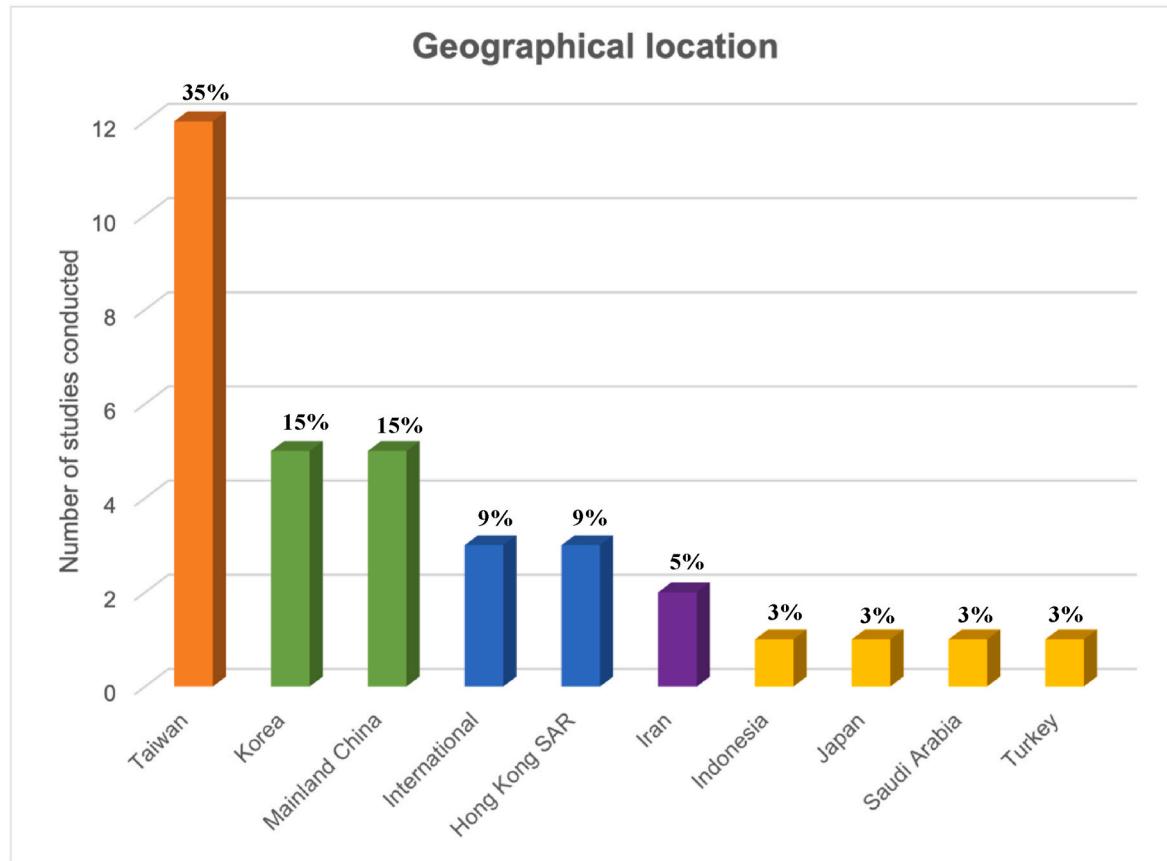


Fig. 5. Publication trend in geographic location.

integrate the research seamlessly into their teaching (Dizon et al., 2020). It offered sampling convenience when the teacher-researchers could invite their classes to join (Yang et al., 2022).

Cross-sectional studies emerged as the second most popular research design at 41%. Researchers often opted for cross-sectional designs to obtain a snapshot of the phenomenon with specific criteria for targeting populations (Levin, 2006). For instance, researchers purposely recruited

fellow educators from his university (Mohamed, 2023) or channels such as ResearchGate for an online discussion (Bin-Hardy et al., 2023). Additionally, cross-sectional design was commonly employed when collecting anonymous data, as in the case of Lin et al.'s study (2023), to gather anonymous responses from parents.

Conversely, the experimental approach was the least popular at 14%. It requires random assignment of participants to minimize potential

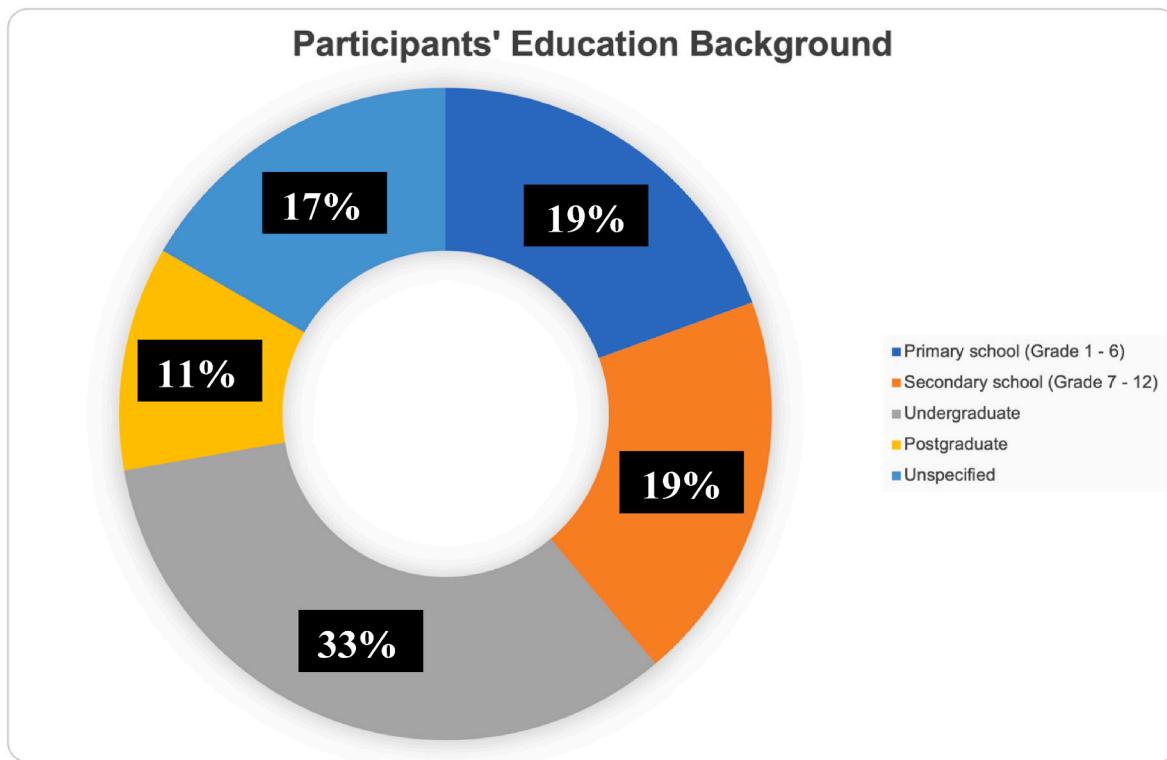


Fig. 6. Publication trend in education background.

biases, which was proven successful with no statistically significant differences in participants' personalities (Um et al., 2023) and oral proficiency in the pre-test (Tai, 2022; Tai & Chen, 2022) to strengthen the validity of results.

4.4. Research trends on data collection and analysis strategies

Fig. 9 shows that 50% of the studies adopted mixed methods, 34% used quantitative methods, and 16% solely collected qualitative data. Among mixed-method studies, Lin and Mubarok (2021) discovered that the speaking pattern was more active with higher word frequency when interacting with AI chatbots through corpus data. Yang et al. (2022) found that 76% of students' utterances in user logs were related to daily life situations, indicating the usefulness of AI chatbots in practicing everyday discourse.

When analyzing conversation logs quantitatively, Yang et al. (2022) suggested that AI chatbots presented more opportunities to speak English with 8.2–11.5 turns. To examine effectiveness in ELT, tests adapted from Cambridge (Feng & Wang, 2023), IELTS (Lin & Mubarok, 2021), listening and speaking assessment (Chien et al., 2022; Dizon, 2020; Tai, 2022) and vocabulary test (Hong et al., 2023) were performed. The data collection and analysis strategies trend presented that empirical studies conducted with AI chatbots have evolved from traditional strategies such as surveys and interviews which depended on students' self-rating to objective assessments based on corpus data, user logs, and international English tests.

4.5. Learning outcomes from conversational AI tools in ELT

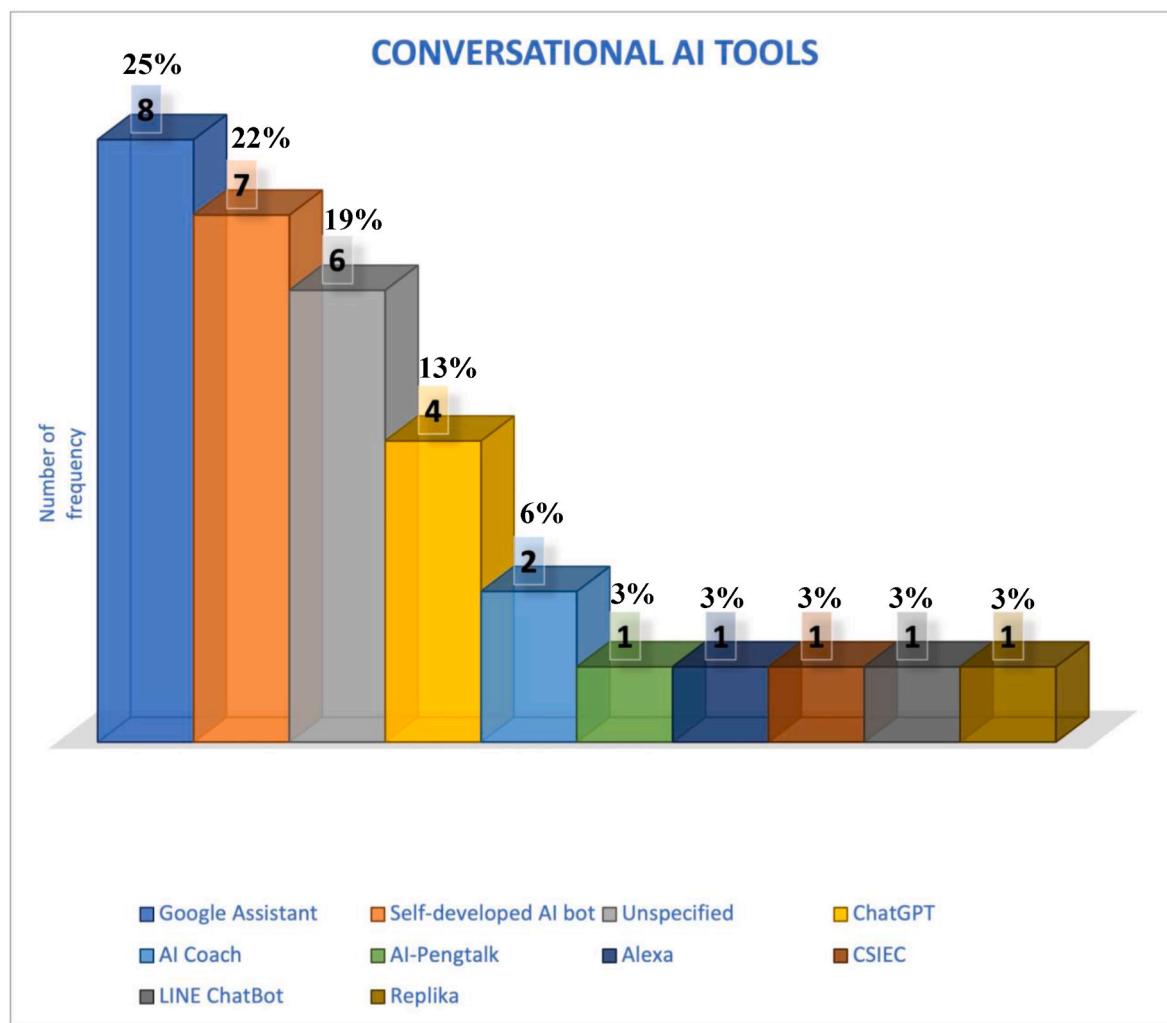
Next, this review presents affective learning outcomes with the highest frequency of 43% ($n = 21$); cognitive outcomes at 41% ($n = 20$), and behavioral outcomes at 16% ($n = 8$) (Appendix D).

Within the affective domain, learners experience enjoyment most ($n = 7$) at 33%, followed by improving learning interest ($n = 5$) at 24%, motivation ($n = 4$) at 19%, lowering foreign language anxiety (FLA) (n

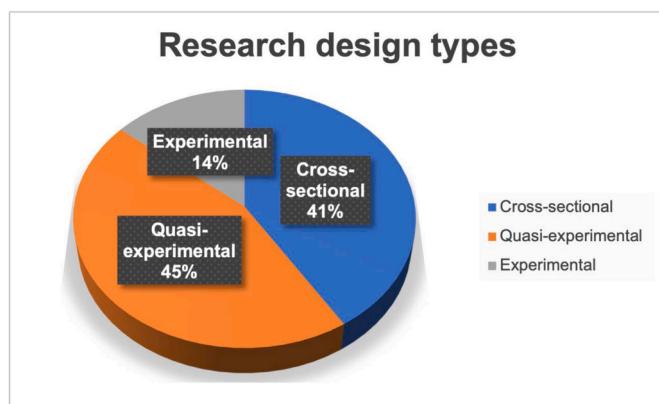
= 2) at 10%, perceived improvement in English ($n = 2$) at 10%, and higher willingness-to-communicate ($n = 1$) at 4%. Students found communicating with the AI chatbots enjoyable (Bin-Hady et al., 2023; Chen et al., 2023), and comfortable chatting in English (Yang et al., 2022), and some enjoyed the supportive and personalized environment (Wang et al., 2022, 2023). Relating to learning interests, most studies suggested AI chatbots enhanced students' interest in ELT specifically (Gonulal, 2023; Tai & Chen, 2023; Um et al., 2023). Moreover, students were more motivated in learning English (Feng & Wang, 2023; Jeon, 2022; Jia et al., 2022; Lin & Mubarok, 2021). Tai and Chen (2022) pointed out that screen-based communication is more effective in lowering FLA than speech-only applications, thus, increasing their willingness to speak in English (Tai & Chen, 2023). The review reflected that even though the main purpose of the studies was to improve targeted ELT outcomes, affective outcomes were achieved by creating an enjoyable and motivating ELT learning environment that sparked students' learning interest and lowered their foreign language anxiety. The review displayed the great potential of enhancing students' learning experience using conversational AI in ELT, in particular, EFL settings.

There were two primary outcomes in the behavioral domain. The highest frequency was the increased opportunities to practice communication ($n = 7$) at 88%. Students gained additional opportunities out-of-class (Tai, 2022; Tai & Chen, 2022), to practice their pragmalinguistic abilities (Timpe-Laughlin et al., 2023), and had in-depth conversations on assigned topics (Lin & Mubarok, 2021). Talking to AI chatbots helped lengthen the time speaking English (Kim et al., 2022; Um et al., 2023), and the number of dialogue exchanges (Yang et al., 2022), which rarely occurred in an EFL classroom. 12% of selected studies reported students increased their reading amount (Feng & Wang, 2023). In the EFL context, researchers reported that there were limited opportunities for students to practice speaking in English, especially outside the classroom (Lee, 2022), the use of conversational AI can create more communication opportunities.

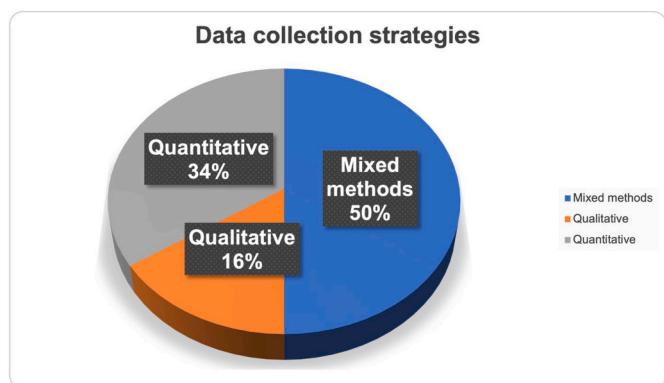
From the scope of the cognitive domain, improvement in speaking skills was most observed ($n = 8$) at 45% from the increased number of

**Fig. 7.** Trends in conversational AI tools.

Notes: AI coach is a voice-driven AI chatbot developed by a Chinese technology company, Arivoc (Wang, Gupta, et al., 2022). AI-Pengtalk is an AI chatbot created by an Education Broading system funded by the South Korean government (Um et al., 2023). CSIEC, Computer Simulation in Educational Communication is a web-based, voice-based interactive AI system (Ebadi & Amini, 2022). Replika is a voice- and text-based AI chatbot accessible on mobile and computers (Lin & Mubarok, 2021).

**Fig. 8.** Types of research design.

proper replies (Hassani et al., 2016), prompts (Bin-Hady et al., 2023), pronunciation (Feng & Wang, 2023), fluency (Kent, 2021; Tai, 2022), less repetition or hesitation (Lin & Mubarok, 2021) and in general (Chien et al., 2022; Dizon, 2020; Tai & Chen, 2022). Subsequently, a

**Fig. 9.** Data collection and analysis strategies.

decrease in grammar errors ($n = 2$), elevated pronunciation ($n = 2$), autonomous learning ($n = 2$), expanded vocabulary ($n = 2$), enhanced attention level ($n = 2$) at 10%, and followed by improvements in comprehension skills ($n = 1$), writing abilities ($n = 1$) and listening skills ($n = 1$) at 5% were observed. Besides linguistic skills, AI chatbots could

help strengthen students' attention levels (Hsu, 2022) and promote autonomous learning by motivating students to seek additional content (Tai, 2022). The results showed that researchers started to shift the focus from language outcomes to explore how AI can improve students' cognitive skills which they can benefit from lifelong learning, such as self-regulated learning (Xia et al., 2023) and critical thinking skills (Darwin et al., 2024).

4.6. Influential ecological factors on L2 learners' usage of conversational AI tools

Following the Ecological System Theory, this review reported from the individual layer at 47% ($n = 34$), microsystem at 30% ($n = 22$), mesosystem at 3% ($n = 2$), exosystem at 11% ($n = 8$), macrosystem at 3% ($n = 2$) to chrono-system factors at 6% ($n = 4$) (Appendix E). Fig. 10 presents a finalized composition of influential factors based on the Ecological System theory.

4.6.1. Individual layer

The individual layer received the utmost attention in affecting students' adaptation of conversational AI tools in ELT, resulting in 47%. Attitude and perception ($n = 18$) were observed to be most influential with over 50% frequency, followed by 32% of non-cognitive factors ($n = 11$), education background ($n = 2$) at 6%, personality ($n = 1$), tech-savviness ($n = 1$) and prior knowledge and skills ($n = 1$) at 3% (Appendix F). Among perceptual factors, students' perception of usefulness has the predominant emphasis at 26% that AI chatbots could improve writing (Escalante et al., 2023), communication (Um et al., 2023; Yang et al., 2022), pronunciation (Gonulal, 2023) and reduce grammatical errors (Hwang et al., 2023). Moreover, students found image recognition (Jia et al., 2022), and multimodality (Tai, 2022) useful in learning English. These findings were depicted from studies using various AI chatbots, including Google Assistant, AI-Pengtalk, ChatGPT, and self-developed bots, which showcased that students' perception of usefulness was not limited to specific bots. Some students perceived their interactions were similar to talking to a native speaker (Gonulal, 2023) and enjoyed social connection from greetings and encouragement (Liu et al., 2022).

In addition, a positive attitude toward the bot (Feng & Wang, 2023; Liu & Ma, 2023) and appreciation toward English (Xia et al., 2023) increased the chatbot usage in ELT. The easiness of use helped sustain

their willingness to communicate with the chatbot (Jeon, 2022). However, Liu and Ma (2023) argued that easiness had to be complemented with other advantages, such as usefulness, to lead to students' adaptation.

Non-cognitive factors, such as reducing FLA in a supportive environment, encouraged the usage of AI chatbots in ELT (Chen, 2022; Tai & Chen, 2022, 2023; Yang et al., 2022). Students found talking to chatbots less stressful (Jeon, 2022) and had a higher level of meditation compared to human partners (Hsu, 2022).

Students' English proficiency affected their adaptation to AI bots in ELT. For instance, students with lower English proficiency found it challenging to speak to the chatbots (Jeon; 2022). Xia et al. (2023) shared that English knowledge is a prerequisite for self-regulated learning with a chatbot. Besides language level, students with higher tech-savviness reacted positively to AI chatbots (Chen, 2022).

4.6.2. Microsystem layer

Individual and microsystem layers were the most influential factors in the selected studies, totaling almost 80%. This review identified two main factors in the Microsystem layer (Appendix G).

Du and Gao (2022) stated that teachers' perception of effectiveness is the most influential, reflected as 18%. Teachers tended to promote AI chatbots when perceived as useful in providing individualized (Escalante et al., 2023), rapid and accurate feedback (Mohamed, 2023), or beneficial in improving fluency (Kent, 2021). Features such as flexibility for students to learn anytime, anywhere (Du & Gao, 2022), supportive (Tai & Chen, 2023), and enjoyable environment (Kent, 2021) contributed to teachers' positive attitude. Teachers' willingness had a significant impact as they might be the first source for students to learn about AI chatbots in ELT (Chen et al., 2023; Jia et al., 2022; Xia et al., 2023).

Factors associated with parents exhibited a dispersed distribution at 32%. Students were more likely to use AI chatbots to learn English if their parents had a positive attitude towards chatbots' practicality, human-like conversation, smart appearance, or personality in accepting new technology (Feng & Wang, 2023; Lin et al., 2023). However, parents might intervene with students' usage because of the increase in screen time (Lin et al., 2023).

4.6.3. Mesosystem layer

The mesosystem and macrosystem layers received the least attention among the selected studies, each at 3% (Appendix H). In comparison to chatbots' usefulness and easiness, school policies such as incentives for teachers were least motivating (Du & Gao, 2022). Additionally, Um et al. (2023) stated that many schools lacked the infrastructure to support online learning during the pandemic, limiting students' exposure to technologies.

4.6.4. Exosystem layer

This review identified three exosystem layer factors: 75% originated from AI tools' advancement ($n = 6$), and 12.5% occurred from parents' background ($n = 1$) and teachers' digital literacy ($n = 1$) each (Appendix I).

Students found that automated speech recognition demotivated them to use AI chatbots to learn English as they struggled to recognize EFL students' accents (Tai, 2022), making them frustrated (Tai & Chen, 2023), anxious, worrisome, exhausted, and eventually, unwilling to speak with chatbots in English (Jeon, 2022). Chen et al. (2023) shared Google Assistant was too sensitive to pronunciation errors and the speed and vocabulary choice were difficult for EFL students. External factors, such as the classroom background noise caused by simultaneous chatbot interactions, hindered the speech recognition capabilities, decreasing students' confidence in speaking English (Kim et al., 2022). Conversely, some students appreciated Google Assistant's ability to generate unique and humorous responses; however, the charm of humor might diminish over time (Gonulal, 2023).

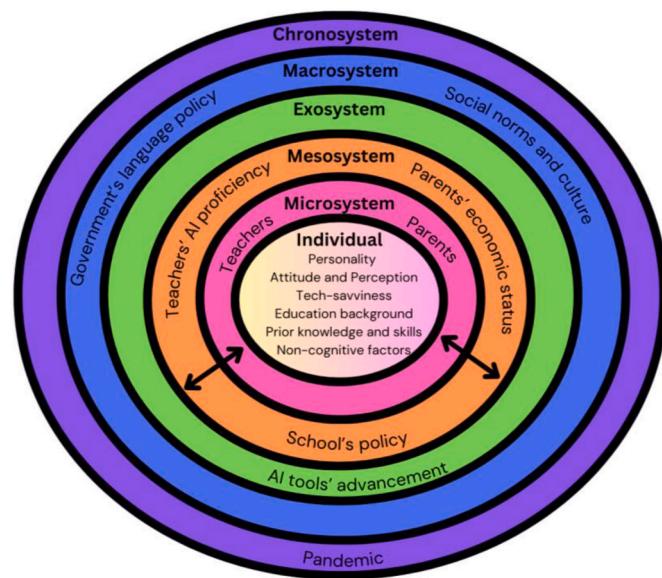


Fig. 10. Finalized adaptation of Ecological System Theory in conversational AI and English context.

Parents' backgrounds and teachers' digital literacy also played an essential role. The gap between parents with lower and higher socio-economic status incurred their access to new technology, especially during the pandemic (Um et al., 2023). Du and Gao (2022) shared that teachers must not find the AI chatbots too complex before introducing them to the students.

4.6.5. Macrosystem layer

The selected studies showed minimal focus at 3% on the macrosystem layer (Appendix J). According to Um et al. (2023), the Korean government encouraged students to use AI chatbots for ELT purposes by developing an AI chatbot 'AI-Pengtalk'. Furthermore, Timpe-Laughlin et al. (2023) highlighted that the social norm of integrating conversational AI tools was inevitable as their functionality extended beyond mere communication today.

4.6.6. Chronosystem layer

The significant change in the chronosystem layer is the COVID-19 pandemic in 2020–2022 (Appendix K). After the pandemic, the mode of communication changed rapidly (Timpe-Laughlin et al., 2023) and shaped the way students learn (Hwang et al., 2023). Jia et al. (2022) asserted a pressing need to achieve authentic language learning through AI-enabled tools, especially in areas that lacked school-based learning opportunities due to suspension. To accommodate the social distancing policies, the government urged the AI chatbot 'AI-Pengtalk' to be introduced earlier than planned, which made it timely and accessible for students to learn English (Um et al., 2023).

5. Implications for practice and research

5.1. Implications for educational research

5.1.1. Growing demand for self-developed chatbots for EFL students

RQ2 findings indicate a growing trend in developing self-made chatbots, despite Google Assistant being the most popular. Given the limitations of existing chatbots in addressing ELT needs and recognizing speech from lower proficiency students (Chen et al., 2023; Jeon, 2022; Kim et al., 2022; Tai, 2022; Tai & Chen, 2023; Yang et al., 2022), researchers and educators are encouraged to create chatbots tailored for EFL students to prevent demotivation and lack of confidence (Chen et al., 2023; Wang, Gupta, et al., 2022). Advances in AI, such as GPT-3 (Hwang et al., 2023) and Google's Dialogueflow (Kim et al., 2022), appear to make designing these tools more accessible.

5.1.2. Longitudinal and experimental research design

RQ3 findings indicate that out of the 32 studies reviewed, 19 of them implemented interventions lasting an average of 8.5 weeks, which is roughly equivalent to 2 months. Nevertheless, the long-term sustainability impact on the students' improvements in English proficiency from AI chatbots remains uncertain (Mohamed, 2023; Um et al., 2023), and it is unclear whether the positive outcomes were influenced by the novelty factor (Tai & Chen, 2023). Thus, future research are advised to conduct longitudinal studies to determine if the benefits of using AI chatbots (e.g., ChatGPT) are sustainable and accurately reflect improvements in students' English abilities, beyond the initial novelty effect (Mohamed, 2023; Tai & Chen, 2023; Um et al., 2023). Current trends in research design for implementing conversational AI tools in ELT tend to favor quasi-experimental and cross-sectional studies. This approach is understandable given the recent introduction of tools like ChatGPT. However, randomized controlled trials are considered the gold standard for evaluating educational practices (Sadoff, 2014; Torgerson & Torgerson, 2007). Researchers should, therefore, design experimental studies to ensure strong internal validity, solid pedagogical insights, and higher applicability in national educational practices (Tai, 2022).

5.1.3. Greater emphasis on perception and attitudes

RQ4 findings show that mixed-method studies were the most common approach, followed by qualitative and quantitative studies. However, limited studies examined perceptions (Mohamed, 2023) and emotional gains of parents and students (Gonulal, 2023; Hong et al., 2023; Lin et al., 2023; Liu & Ma, 2023). Most studies focused on achieving specific ELT goals (Hsu, 2022; Lin & Mubarok, 2021; Wang, Gupta, et al., 2022), but students' perception of the chatbot significantly correlates with their engagement in the target ELT activity (Liu et al., 2022). Therefore, researchers are encouraged to prioritize qualitative or mixed-method studies to understand stakeholders' perceptions, especially teachers and parents, who often directly influence students' usage of conversational AI tools in the ELT context.

5.1.4. Shift research focus to explore behavioral outcomes and self-regulated learning

RQ5 findings indicate that behavioral outcomes are understudied. One study showed AI chatbots help build good reading habits in children (Feng & Wang, 2023), which is crucial for EFL students (Janah et al., 2021). Researchers should explore how AI chatbots can foster behavioral changes, especially in reading habits. In the cognitive domain, studies have focused on linguistic gains but overlooked cognitive skills. Autonomous learning is vital for EFL and adult learners (Milosavljević, 2017; Tuan, 2021), and ChatGPT has potential in promoting self-regulated learning (Bin-Hady et al., 2023). Researchers should investigate how AI chatbots can develop cognitive skills that support English learning (Xia et al., 2023).

5.2. Implications for educational practice

5.2.1. Pedagogical reform with conversational AI tools in the EFL classroom

Based on RQ1 findings, we recommend raising awareness of using AI chatbots in ELT. With the rise in publications, it is crucial for researchers and educators to explore new ELT methods and chatbots, keeping pace with conversational AI advancements. For broader applicability of research results, future research could expand sample sizes (Dizon, 2020; Ebadi & Amini, 2022; Lin et al., 2023; Mohamed, 2023) and include participants with diverse English proficiency levels (Feng & Wang, 2023; Yang et al., 2022), educational backgrounds (Wang et al., 2023), and majors (Tai, 2022). The review showed that all studies were conducted in Asia, highlighting common challenges like limited English communication (Lee, 2022), speaking anxiety (Chou, 2018), writing difficulties (Chou, 2013), exam-focused learning (Wang et al., 2018), and teacher-centered culture (Kweon & Spolsky, 2018). Thus, educators are advised to leverage conversational AI tools to enhance speaking and writing skills. EFL teachers are also encouraged to create activities using accessible AI chatbots to boost communication practice and motivate students with new learning experiences.

5.2.2. Collaborated efforts to promote AI chatbot in ELT from macro to individual levels

Based on RQ 6 findings, ecological systems theory suggests a chain reaction where government policies enhance teachers' AI proficiency, which shapes their attitudes and directly impacts students' use of AI chatbots in the ELT context. There are two suggested actions for policymakers: First, governments should offer training that familiarizes teachers with the latest AI tools (Chiu et al., 2022). Teachers are potentially the first source to introduce AI chatbots to students in ELT (Chen et al., 2023). However, teachers often lack formal AI training (Chiu et al., 2022), and such training can improve their AI literacy and confidence in using AI chatbots. As teachers' perceived usefulness of AI is crucial for adoption (Du & Gao, 2022; Xia et al., 2023), training should highlight how AI chatbots can meet specific ELT goals. Second, governments are encouraged to collaborate with commercial and research institutions to create AI chatbots tailored to local students' needs. For example, Korea's AI-Pengtalk was developed based on common ELT

textbooks, providing a safe learning environment for EFL students (Um et al., 2023).

Teachers should make good use of conversational AI in ELT as a conversational partner (Yang et al., 2022) to supplement ELT activities, with teachers playing a critical role in guiding the use of this technology (Zhao, 2022). Apart from teachers, parents serve as catalysts and share a direct influence on students' decisions. Although parents had a generally positive perception of AI chatbots, their concern about the increased screen time hindered children's usage. Designers and marketers should emphasize the utilitarian and hedonic appeal of AI chatbots in ELT to encourage their acceptance and acknowledge the effectiveness of AI chatbots in ELT (Lin et al., 2023).

Lastly, while most studies did not highlight the importance of macro environment factors, Du and Gao (2022) suggested future research should involve stakeholders including language teachers, administrators, and AI designers to verify the generality of findings and explore their influence. Thus, researchers are advised to examine how broader systems—mesosystem, exosystem, and macrosystem—affect students' adaptation to AI chatbots. By addressing these levels, we can ensure a comprehensive and effective integration of AI chatbots in ELT, benefiting students, teachers, and the broader educational ecosystem.

6. Conclusion and limitations

This review analyzed trends in conversational AI tools in ELT from January 2013 to November 2023, examining 32 papers on publication trends, tool types, research methods, learning outcomes, and influencing factors. Publications on this topic have steadily increased, with a notable surge in 2022 and 2023. All the studies were conducted in Asian EFL contexts, highlighting the region's particular interest in and adoption of these tools. Google Assistant emerged as the most widely used AI chatbot. Research methodologies primarily included quasi-experimental and cross-sectional designs, which helped explore the effectiveness of conversational AI tools. Mixed-method approaches were common, offering a comprehensive view of these tools' impacts. The review found that conversational AI tools positively affect both affective and cognitive skill development, enhancing language learning experiences. However, influencing factors were mainly focused on individual and microsystem levels, such as students, parents, and teachers.

Despite its contributions, this study has three limitations and suggestions for future research: First, the review only included articles from

A&HCI and SSCI databases, potentially excluding relevant studies from conference proceedings, book chapters, and other sources. Given the emerging nature of this field, a broader range of sources could offer deeper insights into the effective use of conversational AI tools in ELT, particularly in the EFL context. Second, with the rapid advancement of conversational AI, new research is expected to proliferate. Conducting annual reviews would help stay current with evolving trends and applications. Lastly, while most studies report positive outcomes, some have noted potential negative effects, such as students finding AI chatbots intimidating and stressful (Randall, 2020; Wang, Gupta, et al., 2022), parental concerns over highly autonomous AI bots (Cocosila et al., 2007, 2009; Faqih, 2022; Lin et al., 2021; Louie et al., 2021), and the need for ethical guidelines in educational settings (Escalante et al., 2023; Kohnke et al., 2023). Future research should take a balanced approach, covering both the positive and negative aspects of using conversational AI tools in ELT.

Statements on open data and ethics

The data used in this systematic review is sourced from publicly available databases and literature sources. No private or confidential data was accessed or used in this review.

Authors' notes

We have no known conflict of interest to disclose.

CRediT authorship contribution statement

Wan Yee Winsy Lai: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Ju Seong Lee:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Interchangeable keywords in the search string

| Artificial Intelligence | Conversational |
|-------------------------|---|
| Intelligent | Bot (Fryer & Carpenter, 2006) Chatbot (Io & Lee, 2017; Yang et al., 2022; McTear, 2022) Conversational agent (Alnefaie et al., 2021; Ayedoun et al., 2015; McTear, 2022; Radziwill & Benton, 2017; Ruan et al., 2019) Conversational application (Borsci et al., 2022) Conversational partner (Belda-Medina & Calvo-Ferrer, 2022) Conversational software (Radziwill & Benton, 2017) Dialogue agent (Ruan et al., 2019) Dialogue system (McTear, 2022; Shawar, 2017) Instant messenger (Hill et al., 2015) Language tutor (Wang & Petrina, 2013) Learning assistant (Zografos & Moussiades, 2023) Personal assistant (McTear, 2022) Robot (Chang et al., 2010) Tutor (Belda-Medina & Calvo-Ferrer, 2022; Zografos & Moussiades, 2023) Agent (Ji et al., 2023) Personal (Ji et al., 2023) |

Appendix B. An overview of the findings from the selected studies

| Study and Year | Participants' background (education and location) | Tool types | ELT goals of the study | Methodology | ELT learning outcomes | Influential factors |
|------------------------------|--|-----------------------|---|---|----------------------------------|--|
| Bin-Hady et al. (2023) | Researchers; international | ChatGPT | Unspecified | Qualitative; Cross-sectional | Affective; Cognitive | Microsystem |
| Chen et al. (2023) | Undergraduate; Taiwan | Google Assistant | Speaking | Mixed-method; Cross-sectional | Affective | Individual; Exosystem |
| Chen (2022) | Undergraduate and postgraduate; Taiwan | Self-developed AI bot | Speaking | Mixed-method; Quasi-experimental | Affective; Cognitive | Individual |
| Chien et al. (2022) | Secondary school; Taiwan | LINE Chatbot | Speaking and listening | Quantitative; Quasi-experimental | Cognitive | Nil |
| Dizon (2020) | Undergraduate; Japan | Alexa | Speaking and listening | Quantitative; Quasi-experimental | Cognitive | Individual |
| Du and Gao (2022) | EFL University educators; China | Unspecified | Teachers' perception | Mixed-method; Cross-sectional | Nil | Microsystem; Exosystem |
| Ebadie and Amini (2022) | Undergraduate; Iran | CSIEC | Learner motivation | Mixed-method; Experimental | Nil | Nil |
| Escalante et al. (2023) | Undergraduate and postgraduate; International | ChatGPT | Writing | Mixed-method; Quasi-experimental | Cognitive | Individual; Microsystem |
| Feng and Wang (2023) | Primary school; Mainland China | Unspecified | Reading and autonomous learning ability | Mixed-method; Cross-sectional | Affective; Behavioral; Cognitive | Individual; Microsystem |
| Gonulal (2023) | Undergraduate; Turkey | Google Assistant | Students' perception | Mixed-method; Cross-sectional | Affective | Individual; Microsystem; Exosystem |
| Hassani et al. (2016) | Postgraduate; Iran | Self-developed bot | Speaking and listening | Quantitative; Quasi-experimental | Cognitive | Nil |
| Hong et al. (2023) | Secondary school; Taiwan | Google Assistant | Vocabulary | Quantitative; Cross-sectional | Cognitive | Nil |
| Hsu (2022) | Undergraduate; Taiwan | Unspecified | Attention level | Quantitative; Cross-sectional | Cognitive | Individual |
| Hwang et al. (2023) | Undergraduate; Indonesia | Self-developed AI bot | Writing | Mixed-method; Quasi-experimental | Cognitive | Individual; Chronosystem |
| Jeon (2022) | Primary school; Korea | Google Assistant | Affordances of chatbots | Qualitative; Quasi-experimental | Affective | Individual; Exosystem |
| Jia et al. (2022) | Undergraduate and Postgraduate; Hong Kong SAR | Self-developed AI bot | Vocabulary and grammar | Mixed-method; Cross-sectional | Affective | Individual; Microsystem; Chronosystem |
| Kent (2021) | University educators; Korea | Google Assistant | Teachers' perception | Qualitative; Quasi-experimental | Affective; Cognitive | Microsystem |
| Kim et al. (2022) | Secondary school; Korea | Self-developed AI bot | Speaking | Quantitative; Quasi-experimental | Behavioral | Exosystem |
| Lin and Mubarok (2021) | Undergraduate; Taiwan | Replika | Speaking | Mixed-method; Cross-sectional | Affective; Behavioral; Cognitive | Nil |
| Lin et al. (2023) | Parents; Taiwan | Unspecified | Parents' perception | Quantitative; Quasi-experimental | Nil | Microsystem |
| Liu et al. (2022) | Primary school; Taiwan | Self-developed AI bot | Reading | Mixed-method; Quasi-experimental | Affective | Individual |
| Liu and Ma (2023) | Unspecified; Mainland China, Hong Kong SAR, Taiwan | ChatGPT | Perception | Quantitative; Cross-sectional | Nil | Individual |
| Mohamed (2023) | EFL University educators; Saudi Arabia | ChatGPT | Teachers' perception | Qualitative; Cross-sectional | Nil | Microsystem |
| Tai (2022) | Undergraduate; Taiwan | Google Assistant | Speaking | Mixed-method; Experimental | Affective; Behavioral; Cognitive | Individual; Exosystem |
| Tai and Chen (2023) | Secondary school; Taiwan | Google Assistant | Speaking | Mixed-method; Quasi-experimental and experimental | Affective | Microsystem |
| Tai and Chen (2022) | Secondary school; Taiwan | Google Assistant | Speaking | Mixed-method; Experimental | Affective; Behavioral; Cognitive | Individual; Mesosystem |
| Timpe-Laughlin et al. (2023) | Undergraduate; International | Unspecified | Speaking | Qualitative; Cross-sectional | Behavioral | Macrosystem; Chronosystem |
| Um et al. (2023) | Primary school; Korea | AI-Pengtalk | Unspecified | Quantitative; Experimental | Affective; Behavioral | Individual; Exosystem; Macrosystem; Chronosystem |
| Wang et al. (2023) | Primary school; Mainland China | AI coach Kouyu | The types of learners | Mixed-method; Quasi-experimental | Affective | Individual |
| Wang, Gupta, et al. (2022) | Primary school; Mainland China | AI coach Kouyu | Students' perception | Quantitative; Quasi-experimental | Affective | Individual |

(continued on next page)

(continued)

| Study and Year | Participants' background (education and location) | Tool types | ELT goals of the study | Methodology | ELT learning outcomes | Influential factors |
|--------------------|---|-----------------------|---|----------------------------------|-----------------------|-------------------------|
| Xia et al. (2023) | Secondary school; Hong Kong SAR | Unspecified | Relationship among English, AI literacy and autonomous learning | Quantitative; Cross-sectional | Nil | Individual; Microsystem |
| Yang et al. (2022) | Primary and secondary schools; Korea | Self-developed AI bot | Speaking | Mixed-method; Quasi-experimental | Affective; Behavioral | Individual |

Appendix C. Coding framework for the chosen studies

| Research questions | Themes | Sub-themes | Details | References |
|--------------------|----------------------------------|---|--|---|
| 1 | Trends - Publication | Year | The number of studies published on SSCI or A&HCI journals on a yearly basis from January 2013 to November 2023 | Ji et al. (2023), Sharadgah and Sa'di (2022) |
| | | Journal titles | The frequency of selected articles published in a particular journal | Chen et al. (2021) |
| | | Geographic location | The frequency of countries (e.g., Indonesia) or regions (e.g., Hong Kong) where the selected studies were conducted and presented in percentages | Zhai and Wibowo (2023), Liang et al. (2023), Sharadgah and Sa'di (2022) |
| | | Participants' education background | The frequency and percentage of the participants' education level distribution, including (1) primary education, (2) secondary education, (3) undergraduate, (4) postgraduate, (5) unspecified | Klimova et al. (2023), Sharadgah and Sa'di (2022) |
| 2 | Trends - conversational AI tools | The tools' popularity | The frequency of the conversational AI tools implemented | |
| 3 | Research trends on methodology | Research design types | The methodology in the research design of the selected studies is categorized as (1) cross-sectional, (2) quasi-experimental and (3) experimental. | |
| 4 | | Data collection and analysis strategies | The research collection strategies classified as purely (1) quantitative, (2) qualitative or (3) mixed-method. | Klimova et al. (2023), Liang et al. (2023), Zhai and Wibowo (2023) |
| 5 | Outcomes | The learning outcomes | The learning outcomes from the selected studies, categorized by Affective, Behavioral and Cognitive dimensions following a revision of Bloom's Taxonomy | Krathwohl (2002); Krathwohl et al. (1964) |
| 6 | Influential factors | Influential factors behind the students' adaption of the AI tools | The influential factors are identified based on the six layers of the Bronfenbrenner's ecological system theory from (1) individual, (2) microsystem, (3) mesosystem, (4) exosystem, (5) macrosystem and (6) chronosystem. | Bronfenbrenner (1974, 1979) |

Appendix D. Learning outcomes in affective, behavioral, and cognitive domains

| Categorization | Factors | Frequency | Total N |
|----------------|---|-----------|---------|
| Affective | Enjoyment | 7 | 21 |
| | Learning interest | 5 | |
| | Motivation | 4 | |
| | Lower foreign language anxiety | 2 | |
| | Perceived improvement | 2 | |
| | Higher willingness-to-communicate | 1 | |
| Behavioral | Increased opportunities to practice communication | 7 | 8 |
| | Increased reading amount | 1 | |
| Cognitive | Improved speaking skills | 9 | 20 |
| | Strengthen autonomous learning | 2 | |
| | Decreased grammar errors | 2 | |
| | Expanded in vocabulary usage | 2 | |
| | Enhanced attention level | 2 | |
| | Improved comprehension skills | 1 | |
| | Improved writing abilities | 1 | |
| | Improved listening skills | 1 | |

Note: Frequency = The number of papers mentioning the learning outcomes.

Appendix E. Overview of Ecological System Theory layers and frequency

| Ecological System Theory layers (Total) | Frequency (N = 72) |
|---|--------------------|
| Individual layer | 34 (47%) |
| Microsystem | 22 (30%) |
| Mesosystem | 2 (3%) |
| Exosystem | 8 (11%) |

(continued on next page)

(continued)

| Ecological System Theory layers (Total) | Frequency (N = 72) |
|---|--------------------|
| Macrosystem | 2 (3%) |
| Chronosystem | 4 (6%) |

Appendix F. Individual layer factors

| Categorization | Factors | Frequency (N = 34) |
|------------------------------------|---|---|
| Attitude and Perception (n = 18) | Perception - usefulness Perception - human-like Attitude toward the tool Perception - ease of use Attitude toward English Perception - tool appearance | 9 (26%) 3 (9%) 2 (6%) 2 (6%) 1 (3%) 1 (3%) |
| Non-cognitive factors (n = 11) | Foreign language anxiety (FLA) Motivation Interest Enjoyment Learner autonomy | 6 (17%) 2 (6%) 1 (3%) 1 (3%) 1 (3%) |
| Education background (n = 2) | English proficiency | 2 (6%) |
| Personality (n = 1) | Curiosity | 1 (3%) |
| Tech-savviness (n = 1) | Technology acceptance | 1 (3%) |
| Prior knowledge and skills (n = 1) | Prior learning skills | 1 (3%) |

Note: Frequency = The number of papers mentioning individual factors.

Appendix G. Microsystem layer factors

| Categorization | Factors | Frequency (N = 22) |
|-------------------|--|--|
| Teachers (n = 15) | Perception - usefulness Attitude toward the tool Perception - ease of use Willingness Perception - negative consequences Teaching style | 4 (18%) 3 (14%) 3 (14%) 3 (14%) 1 (4.5%) 1 (4.5%) |
| Parents (n = 7) | Attitude toward the tool Personality Perception - usefulness Perception - human-like Perception - appearance Perception - negative consequences | 2 (9%) 1 (4.5%) 1 (4.5%) 1 (4.5%) 1 (4.5%) 1 (4.5%) |

Note: Frequency = The number of papers mentioning microsystem factors.

Appendix H. Mesosystem layer factors

| Categorization | Factors | Frequency (N = 2) |
|-----------------|-------------------------|-------------------|
| Schools (n = 2) | Schools' infrastructure | 1 (50%) |
| | Incentives for teachers | 1 (50%) |

Note: Frequency = The number of papers mentioning mesosystem factors.

Appendix I. Exosystem layer factors

| Categorization | Factors | Frequency (N = 8) |
|------------------------------------|---|------------------------|
| AI tools' advancement (n = 6) | Automated Speech recognition Humor in language | 5 (62.5%) 1 (12.5%) |
| Parents' background (n = 1) | Social economic status | 1 (12.5%) |
| Teacher's digital literacy (n = 1) | AI proficiency | 1 (12.5%) |

Note: Frequency = The number of papers mentioning exosystem factors.

Appendix J. Macrosystem layer factors

| Categorization | Factors | Frequency (N = 2) |
|--------------------------------------|-----------------------|-------------------|
| Government's language policy (n = 1) | Implementation for AI | 1 (50%) |
| Social norm (n = 1) | Commonality of AI | 1 (50%) |

Note: Frequency = The number of papers mentioning macrosystem factors.

Appendix K. Chronosystem layer factors

| Categorization | Factors | Frequency (N = 4) |
|------------------|---------------------------------------|-------------------|
| Pandemic (n = 4) | Necessity of AI | 2 (50%) |
| | Changes in learning and communication | 2 (50%) |

Note: Frequency = The number of papers mentioning chronosystem factors.

References

- Articles with an asterisk (*) were selected studies for the review.
- Al-Kaisi, A. N., Arkhangelskaya, A. L., & Rudenko-Morgun, O. I. (2019). Voice assistants as a training tool in a foreign language class [Paper presentation]. In *INTED2019-13th international technology, education and development conference proceedings* (pp. 1236–1246). <https://doi.org/10.21125/inted.2019.0399>. Valencia.
- Alharbi, W. (2023). AI in the foreign language classroom: A pedagogical overview of automated writing assistance tools. *Educational Research International*, 1–15. <https://doi.org/10.1155/2023/425331>
- Alnefaie, A., Singh, S., Kocaballi, B., & Prasad, M. (2021). An overview of conversational agent: Applications, challenges and future directions. In *17th international conference on web information systems and technologies*. SCITEPRESS-Science and Technology Publications.
- Ayedoun, E., Hayashi, Y., & Seta, K. (2015). A conversational agent to encourage willingness to communicate in the context of English as a foreign language. *Procedia Computer Science*, 60, 1433–1442.
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in Language Learning. *Applied Sciences*, 12(17), 8427. <https://doi.org/10.3390/app12178427>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? □. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623). <https://doi.org/10.1145/3442188.3445922>
- * Bin-Hady, W. R. A., Al-Kadi, A., Hazaera, A., & Ali, J. K. M. (2023). Exploring the dimensions of ChatGPT in English language learning: A global perspective. *Library Hi Tech*. <https://doi.org/10.1108/LHT-05-2023-0200>.
- Borsci, S., Malizia, A., Schmettow, M., Van Der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2022). The chatbot usability scale: The design and pilot of a usability scale for interaction with AI-based conversational agents. *Personal and Ubiquitous Computing*, 26, 95–119.
- Bronfenbrenner, U. (1974). Is early intervention effective? *Teachers College Record*, 76(2), 1–19.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard university press.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Chang, C. K., Siagian, C., & Itti, L. (2010). Mobile robot vision navigation & localization using gist and saliency. In *2010 IEEE/RSJ international Conference on intelligent Robots and systems* (pp. 4147–4154). IEEE.
- Chen, X., Zou, D., Cheng, K. S., & Xie, H. (2021). Artificial intelligence-assisted personalized language learning: Systematic review and co-citation analysis. In *Artificial intelligence-assisted personalized language learning: Systematic review and co-citation analysis*. IEEE.
- * Chen, Y. C. (2022). Effects of technology-enhanced language learning on reducing EFL learners' public speaking anxiety. *Computer Assisted Language Learning*, 1–25.
- Chen, H. L., Vicki Widarso, G., & Sutrisno, H. (2020). A ChatBot for learning Chinese: Learning achievement and technology acceptance. *Journal of Educational Computing Research*, 58(6), 1161–1189. <https://doi.org/10.1177/0735633120929622>
- Chen, H. H.-J., Yang, C. T.-Y., & Lai, K. K.-W. (2023). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments*, 31(3), 1335–1350. <https://doi.org/10.1080/10494820.2020.1833043>
- * Chen, H. H. J., Yang, C. T. Y., & Lai, K. K. W. (2023). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments*, 31(3), 1335–1350.
- * Chien, Y. C., Wu, T. T., Lai, C. H., & Huang, Y. M. (2022). Investigation of the influence of artificial intelligence markup language-based LINE chatbot in contextual English learning. *Frontiers in Psychology*, 13, Article 785752. <https://doi.org/10.3389/fpsyg.2022.785752>.
- Chiu, T. K. F., Meng, H., Chai, C.-S., King, I., Wong, S., & Yam, Y. (2022). Creation and evaluation of a pretertiary artificial intelligence (AI) curriculum. *IEEE Transactions on Education*, 65(1), 30–39. <https://doi.org/10.1109/TE.2021.3085878>
- Chocarro, R., Cortiñas, M., & Marcos-Matás, G. (2023). Teachers' attitudes towards chatbots in education: A technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics. *Educational Studies*, 49(2), 295–313. <https://doi.org/10.1080/03055698.2020.1850426>
- Chou, S. (2013). Language anxiety in second language writing: Is it really a stumbling block? *Second Language Studies*, 31(2), 1–42.
- Chou, M. (2018). Speaking anxiety and strategy use for learning English as a Foreign Language in full and partial English-medium instruction contexts. *Tesol Quarterly*, 52 (3), 611–633. <https://doi.org/10.1002/tesq.455>
- Cocosila, M., Archer, N., & Yuan, Y. (2009). Early investigation of new information technology acceptance: A perceived risk - motivation model. *Communications of the Association for Information Systems*, 25, 30. <https://doi.org/10.17705/1CAIS.02530>
- Cocosila, M., Turel, O., Archer, N., & Yuan, Y. (2007). Perceived health risks of 3G cell phones. *Communications of the ACM*, 50(6), 89–92. <https://doi.org/10.1145/1247001.1247026>
- Creswell, J. W. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.
- Darwin, Rusdin, D., Mukminatin, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education*, 11(1). <https://doi.org/10.1080/23311>
- * Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language, Learning and Technology*, 24(1), 16, 10.125/44705.
- * Du, Y., & Gao, H. (2022). Determinants affecting teachers' adoption of AI-based applications in EFL context: An analysis of analytic hierarchy process. *Education and Information Technologies*, 27(7), 9357–9384. <https://doi.org/10.1007/s10639-022-11001-y>.
- * Ebadi, S., & Amini, A. (2022). Examining the roles of social presence and human-likeness on Iranian EFL learners' motivation using artificial intelligence technology: A case of CSIEC chatbot. *Interactive Learning Environments*, 1–19. <https://doi.org/10.1080/10494820.2022.2096638>.
- * Escalante, J., Pack, A., & Barrett, A. (2023). AI-Generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00425-2>, 57–20.
- Faqih, K. M. S. (2022). Factors influencing the behavioral intention to adopt a technological innovation from a developing country context: The case of mobile augmented reality games. *Technology in Society*, 69, Article 101958. <https://doi.org/10.1016/j.techsoc.2022.101958>
- * Feng, Y., & Wang, X. (2023). A comparative study on the development of Chinese and English abilities of Chinese primary school students through two bilingual reading modes: Human-AI robot interaction and paper books. *Frontiers in Psychology*, 14, Article 1200675. <https://doi.org/10.3389/fpsyg.2023.1200675>.
- Fitriya, T. N. (2021). The use technology based on artificial intelligence in English teaching and learning. *ELT Echo: The Journal of English Language Teaching in Foreign Language Context*, 6(2), 213–223. <https://doi.org/10.24235/eltecho.v6i2.9299>
- Fryer, L., & Carpenter, R. (2006). Bots as Language Learning tools. *Language, Learning and Technology*, 10(3), 8.
- * Gonulal, T. (2023). Investigating EFL learners' humorous interactions with an intelligent personal assistant. *Interactive Learning Environments*, 31(7), 4521–4534. <https://doi.org/10.1080/10494820.2021.1974489>.
- Guo, X., & Lee, J. S. (2023). A systematic review of Informal Digital Learning of English: An ecological systems theory perspective. *System*, 117, Article 103097. <https://doi.org/10.1016/j.system.2023.103097>
- * Hassani, K., Nahvi, A., & Ahmadi, A. (2016). Design and implementation of an intelligent virtual environment for improving speaking and listening skills. *Interactive Learning Environments*, 24(1), 252–271. <https://doi.org/10.1080/10494820.2013.846265>.
- Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Computers in Human Behavior*, 49, 245–250.
- * Hong, J.-C., Lin, C.-H., & Juh, C.-C. (2023). Using a chatbot to learn English via charades: The correlates between social presence, hedonic value, perceived value, and learning outcome. *Interactive Learning Environments*, 1–17. <https://doi.org/10.1080/10494820.2023.2273485>.
- * Hsu, L. (2022). To CALL or not to CALL: Empirical evidence from neuroscience. *Computer Assisted Language Learning*, 35(4), 792–815. <https://doi.org/10.1080/09588221.2020.1750429>.
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—art they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), 237–257. <https://doi.org/10.1111/jcal.12610>
- * Hwang, W.-Y., Nurtantyana, R., Purba, S. W. D., Hariyanti, U., Indrihapsari, Y., & Surjono, H. D. (2023). AI and recognition technologies to facilitate English as foreign language writing for supporting personalization and contextualization in authentic contexts. *Journal of Educational Computing Research*, 61(5), 1008–1035. <https://doi.org/10.1177/07356331221137253>.
- Io, H. N., & Lee, C. B. (2017). Chatbots and conversational agents: A bibliometric analysis. In *2017 IEEE international conference on industrial engineering and engineering management (IEEM)* (pp. 215–219). <https://doi.org/10.1109/IEEM.2017.8289883>
- Janah, K. E. N., Latief, M. A., & Andreani, S. (2021). Exploring EFL students' English reading habits: Reading beyond classroom practice. *JET (Journal of English Teaching Adi Buana)*, 6(1), 17–31.

- * Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, 1–26. <https://doi.org/10.1080/09588221.2021.2021241>.
- Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, 55(1), 48–63. <https://doi.org/10.1080/15391523.2022.2142873>
- * Jia, F., Sun, D., Ma, Q., & Looi, C. (2022). Developing an AI-based learning system for L2 learners' authentic and ubiquitous learning in English language. *Sustainability*, 14, Article 15527. <https://doi.org/10.3390/su142315527>.
- Jiao, W., Wang, W., Jen-tse, H., Wang, X., Shi, S., & Tu, Z. (2023). Is ChatGPT A good translator? Yes with GPT-4 as the engine. <https://doi.org/10.48550/arxiv.2301.08745>. arXiv.Org.
- * Kent, D. (2021). Voice-user interfaces for TESOL: Potential and receptiveness among native and non-native English speaking instructors. *Language, Learning and Technology*, 25(3), 27.
- Kim, H., Shin, D. K., Yang, H., & Lee, J. H. (2019). A study of AI chatbot as an assistant tool for school English curriculum. *Journal of Learner-Centered Curriculum and Instruction*, 19(1), 89–110. <https://doi.org/10.22251/jlcci.2019.19.1.89>
- * Kim, H., Yang, H., Shin, D., & Lee, J. H. (2022). Design principles and architecture of a second language learning chatbot. *Language Learning*, 26(1), 1–18.
- Klimova, B., Pikhart, M., Polakova, P., Cerna, M., Yayilgan, S. Y., & Shaikh, S. (2023). A systematic review on the use of emerging technologies in teaching English as an applied language at the university level. *Systems*, 11(1), 42. <https://doi.org/10.3390/systems11010042>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for Language Teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/0036882231162868>
- Krathwohl, D. R. (2002). *A Revision of Bloom's Taxonomy: An Overview. Theory Into Practice*, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals, Handbook II: Affective domain*. New York: David McKay Company Incorporated.
- Kruk, M., Pawlak, M., Shirvan, M. E., & Shahnama, M. (2022). The emergence of boredom in an online language class: An ecological perspective. *System*, 107, 1–16. <https://doi.org/10.1016/j.system.2022.102803>
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973–1018.
- Kulkarni, P., Mahabaleshwarkar, A., Kulkarni, M., Sirsikar, N., & Gadgil, K. (2019). Conversational AI: An overview of methodologies, applications & future scope. In *2019 5th international conference on computing, communication, control and automation (ICCUBEA)* (pp. 1–7). <https://doi.org/10.1109/ICCUBEA47591.2019.9129347>
- Kweon, S. O., & Spolsky, B. (Eds.). (2018). *The asian EFL classroom: Issues, challenges and future expectations*. Routledge.
- Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 6, Article 100174. <https://doi.org/10.1016/j.caeo.2024.100174>
- Lee, J. S. (2022). *Informal digital learning of English: Research to practice*. Routledge.
- Levin, K. (2006). Study design III: Cross-sectional studies. *Evidence-Based Dentistry*, 7(1), 24–25. <https://doi.org/10.1038/sj.ebd.6400375>
- Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French/y. *CALICO Journal*, 32(1), 1–25. <https://doi.org/10.1558/cj.v32i1.25962>
- Liang, J.-C., Hwang, G.-J., Chen, M.-R. A., & Darmawansah, D. (2023). Roles and research foci of artificial intelligence in language education: An integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*, 31(7), 4270–4296. <https://doi.org/10.1080/10494820.2021.1958348>
- Lin, M. P. C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot: A mixedmethod classroom study. *Journal of Educational Technology & Society*, 23(1), 78–92.
- * Lin, G.-Y., Jhang, C.-C., & Wang, Y.-S. (2023). Factors affecting parental intention to use AI-based social robots for children's ESL learning. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-12023-w>.
- * Lin, C. J., & Mubarok, 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.
- Lin, C., Šabanović, S., Dombrowski, L., Miller, A. D., Brady, E., & MacDorman, K. F. (2021). Parental acceptance of children's storytelling robots: A projection of the uncanny valley of AI. *Frontiers in Robotics and AI*, 8. <https://doi.org/10.3389/frobt.2021.579993>, 579993–579993.
- Liub, B. (2023). Chinese university students' attitudes and perceptions in learning English using ChatGPT. *International Journal of Education and Humanities*, 3(2), 132–140. <https://doi.org/10.58557/jieh.v3i2.145>
- * Liu, C.-C., Liao, M.-G., Chang, C.-H., & Lin, H.-M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education*, 189, Article 104576. <https://doi.org/10.1016/j.compedu.2022.104576>.
- * Liu, G., & Ma, C. (2023). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, 1–14. <https://doi.org/10.1080/17501229.2023.2240316>.
- Louie, B., Björling, E. A., & Kuo, A. C. (2021). The desire for social robots to support English language learners: Exploring robot perceptions of teachers, parents, and students. *Frontiers in Education (Lausanne)*, 6. <https://doi.org/10.3389/feduc.2021.566909>
- Mackey, A., & Gass, S. M. (2016). In *Second language research : Methodology and design* (2nd ed.). Routledge.
- McTear, M. (2022). *Conversational ai: Dialogue systems, conversational agents, and chatbots*. Springer Nature.
- Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. In *FinTech and artificial intelligence for sustainable development: The role of smart technologies in achieving development goals* (pp. 387–409). Cham: Springer Nature Switzerland.
- Milosavljević, N. (2017). Importance of autonomy in adult learners of English for specific purposes. *Journal of Teaching English for Specific and Academic Purposes*, 5(3), 509. <https://doi.org/10.22190/JTESAP1703509M>
- * Mohamed, A. M. (2023). 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*. <https://doi.org/10.1007/s10639-023-11917-z>.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), Article e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Nghi, T., Huu Phuc, T., & Nguyen Tat, T. (2019). Applying AI chatbot for teaching a foreign language: An empirical research. *International Journal of Scientific & Technology Research*, 8(12), 897–902.
- Nguyen, T. T. H. (2023). EFL teachers' perspectives toward the use of ChatGPT in writing classes: A case study at van lang university. *International Journal of Legal Information*, 2(3), 1–47. <https://doi.org/10.54855/ijli.23231>
- Ou, A. W., Stöhr, C., & Malmström, H. (2024). Academic communication with AI-powered language tools in higher education: From a post-humanist perspective. *System*, 121, Article 103225. <https://doi.org/10.1016/j.system.2024.103225>
- Qadir, J. (2023). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In *2023 IEEE global engineering education conference (EDUCON)* (pp. 1–9). IEEE.
- Qasem, F., Ghaleb, M., Mahdi, H. S., Khateeb, A. A., & Fadda, A. H. (2023). Dialog chatbot as an interactive online tool in enhancing ESP vocabulary learning. *Saudi Journal of Language Studies*. <https://doi.org/10.1108/SJLS-10-2022-0072>
- Radziwill, N. M., & Benton, M. C. (2017). *Evaluating quality of chatbots and intelligent conversational agents*. <https://doi.org/10.48550/arxiv.1704.04579>. arXiv.Org.
- Randall, N. (2020). A survey of robot-assisted Language Learning (RALL). *ACM Transactions on Human-Robot Interaction*, 9(1), 1–36. <https://doi.org/10.1145/3345506>
- Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and open-AI models: A preliminary review. *Future Internet*, 15(6), 192. <https://doi.org/10.3390/fi15060192>
- Ruan, S., Jiang, L., Xu, Q., Liu, Z., Davis, G. M., Brunsell, E., & Landay, J. A. (2021). EnglishBot: An AI-powered conversational system for second Language Learning. In *26th international conference on intelligent user interfaces* (pp. 434–444). <https://doi.org/10.1145/3397481.3450648>
- Ruan, S., Jiang, L., Xu, J., Tham, B. J. K., Qiu, Z., Zhu, Y., ... Landay, J. A. (2019). Quizbot: A dialogue-based adaptive learning system for factual knowledge. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–13).
- Sadoff, S. (2014). The role of experimentation in education policy. *Oxford Review of Economic Policy*, 30(4), 597–620.
- Shalevskaya, E. (2023). AI language models, standardized tests, and academic integrity: A chat (gpt). *International journal of Education Teacher*, 26, 17–25.
- Sharadgah, T., & Sa'di, R. (2022). A systematic review of research on the use of artificial intelligence in English Language Teaching and learning (2015–2021): What are the current effects? *Journal of Information Technology Education: Research*, 21, 337–377. <https://doi.org/10.28945/4999>
- Shawar, B. A. (2017). Integrating CALL systems with chatbots as conversational partners. *Computación Y Sistemas*, 21(4), 615–626.
- Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the facebook messenger. *Computers & Education*, 151, Article 103862. <https://doi.org/10.1016/j.compedu.2020.103862>
- Snowman, J., McCown, R., & Beihler, R. (2011). In *Psychology applied to teaching* (13th ed.). Canada: Wardworth Cengage Learning.
- * Tai, T.-Y. (2022). Effects of intelligent personal assistants on EFL learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*, 1–30. <https://doi.org/10.1080/09588221.2022.2075013>
- * Tai, T.-Y., & Chen, H. H.-J. (2022). The impact of intelligent personal assistants on adolescent EFL learners' speaking proficiency. *Computer Assisted Language Learning*, 1–28. <https://doi.org/10.1080/09588221.2022.2070219>.
- * Tai, T.-Y., & Chen, H. H.-J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, 31(3), 1485–1502. <https://doi.org/10.1080/10494820.2020.1841801>.
- Thompson, C. B., & Panacek, E. A. (2007). Research study designs: Non-experimental. *Air Medical Journal*, 26(1), 18–22.
- * Timpe-Laughlin, V., Dombi, J., Sydorenko, T., & Sasayama, S. (2023). L2 learners' pragmatic output in a face-to-face vs. a computer-guided role-play task: Implications for TBLT. *Language Teaching Research*, , Article 13621688231188310. <https://doi.org/10.1177/13621688231188310>.
- Torgerson, C. J., & Torgerson, D. J. (2007). The need for pragmatic experimentation in educational research. *Economics of Innovation and New Technology*, 16(5), 323–330.
- Tuan, D. M. (2021). Learner autonomy in English Language Learning: Vietnamese EFL learners' perceptions and practices. *Indonesian Journal of Applied Linguistics*, 11(2). <https://doi.org/10.17509/ijal.v1i12.29605>
- Ulla, M. B., Perales, W. F., & Busbus, S. O. (2023). To generate or stop generating response: Exploring EFL teachers' perspectives on ChatGPT in English language

- teaching in Thailand. *Learning: Research and Practice*, 1–15. <https://doi.org/10.1080/23735082.2023.2257252>
- * Um, H., Kim, H., Choi, D., & Oh, H. (2023). An AI-based English education platform during the COVID-19 pandemic. *Universal Access in the Information Society*. <https://doi.org/10.1007/s10209-023-01046-2>.
- Wang, H., Gupta, S., Singhal, A., Muttreja, P., Singh, S., Sharma, P., & Piterova, A. (2022). An artificial intelligence chatbot for young people's sexual and reproductive health in India (SnehAI): Instrumental case study. *Journal of Medical Internet Research*, 24(1), Article e29969. <https://doi.org/10.2196/29969>
- * Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers and Education*, 194, Article 104703. <https://doi.org/10.1016/j.compedu.2022.104703>.
- * Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2022). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*, 1–27. <https://doi.org/10.1080/09588221.2022.2056203>.
- Wang, Y. F., & Petrina, S. (2013). Using learning analytics to understand the design of an intelligent language tutor—Chatbot Lucy. *Editorial Preface*, 4(11), 124–131.
- * Xia, Q., Chiu, T. K. F., Chai, C. S., & Xie, K. (2023). The mediating effects of needs satisfaction on the relationships between prior knowledge and self-regulated learning through artificial intelligence chatbot. *British Journal of Educational Technology*, 54(4), 967–986. <https://doi.org/10.1111/bjet.13305>.
- * Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, 34(3), 327–343. <https://doi.org/10.1017/S0958344022000039>.
- Yu, H., & Nazir, S. (2021). Role of 5g and artificial intelligence for research and transformation of English situational teaching in higher studies. *Mobile Information Systems*, 2021, 1–16. <https://doi.org/10.1155/2021/3773414>
- Zhai, C., & Wibowo, S. (2022). A systematic review on cross-culture, humor and empathy dimensions in conversational chatbots: The case of second language acquisition. *Heliyon*, 8(12), Article e12056. <https://doi.org/10.1016/j.heliyon.2022.e12056>
- Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence*, 4, Article 100134. <https://doi.org/10.1016/j.caeari.2023.100134>
- Zhao, X. (2022). Leveraging artificial intelligence (AI) technology for English writing: Introducing wordtune as a digital writing assistant for EFL writers. *RELC Journal*, Article 00336882221094089.
- Zografos, G., & Moussiades, L. (2023). A gpt-based vocabulary tutor. In *International conference on intelligent tutoring systems* (pp. 270–280). Cham: Springer Nature Switzerland.