Exploring the Effects of an AI Chatbot on Emotional Engagement in English Speaking Lessons: Insights from Call Annie

Nguyen Lam Anh Duong[1], Le Thi Thien Phuoc[1*]

[1]Industrial University of Ho Chi Minh City, Vietnam
[*]Corresponding author's email: lethithienphuoc@iuh.edu.vn
[*] https://orcid.org/0000-0001-6168-1127
https://doi.org/10.54855/ijaile.25225

Received: 02/05/2025 Revision: 08/07/2025 Accepted: 14/07/2025 Online: 15/07/2025

ABSTRACT

The use of artificial intelligence (AI) in English language teaching is becoming more common, especially in enhancing students' confidence levels in speaking lessons. Existing literature has provided insights into how AI can be incorporated to facilitate students' learning in diverse contexts, although the number of studies on how AI chatbots can improve students' emotional engagement is limited. This study aims to explore how using an AI tool called Call Annie affects students' emotional and behavioral engagement when assigned as homework. 77 English-majored students participated in a three-week project where they used the application to have regular conversations with Call Annie – an AI-powered conversational agent. Two questionnaires were administered in the first and third weeks. Emotional engagement levels, regarding confidence and anxiety, concerns, and behavioral degree related to willingness to communicate (WTC) were compared using a paired-sample T-Test. The results indicated increased levels of confidence and WTC, despite some ongoing concerns about the use of the tool. These findings suggest that Call Annie can enhance emotional engagement, paving the way for more personalized learning experiences through the use of AI chatbots.

Keywords:
Generative AI, students' engagement, virtual conversational partner, willingness to communicate

Introduction

Since the introduction of Chat GPT in 2022, Generative Artificial Intelligence (GAI) has presented significant opportunities for both scholars and teachers to explore the potential benefits of GenAI on language learning (Yu & Guo, 2023), particularly in terms of emotions and learners' motivation. From a cultural perspective, factors that inhibit verbal performance may be influenced by a fear of losing face or shyness (Bao, 2019). Thanks to the development of GAI, factors regarding students' motivation in learning have been expanded in numerous studies, including students' engagement in studying a second language. This has led L2 researchers to examine the impact of AI-powered tools and chatbots on enhancing student

CITATION | Nguyen, L. A. D., & Le, T. T. P. (2025). Exploring the Effects of an AI Chatbot on Emotional Engagement in English Speaking Lessons: Insights from Call Annie. *International Journal of AI in Language Education, 2*(2), 79-99. DOI: https://doi.org/10.54855/ijaile.25225

academic involvement (Nguyen, 2024; Rad, Alipour, & Jafarpour, 2023; Xu, Dugdale, Wei, & Mi, 2023).

As a remedy for students' shyness and reluctance during conversations in English, AI conversational partners might serve as an aid, offering private and non-judgmental practice initiatives. To be more specific, a lower level of anxiety and the enhancement of confidence levels are what could be observed for those practicing with AI chatbots, as stated by Ramos et al. (2022). This anxiety reduction can be especially beneficial for students from cultures where public performance and accuracy are highly emphasized. Besides emotional engagement, willingness to communicate (WTC) could be associated with the degree to which students had a choice to participate in conversations. This layer of motivation was highlighted by MacIntyre et al. (1998), who selected appropriate teaching methods to suit individualized learning styles in class.

Besides the favoured features and use of AI chatbots in second language learning, there have been a number of drawbacks associated with this adoption. One major issue reported by several researchers is the accuracy and cultural sensitivity of the answers generated by AI applications. Golonka et al. (2014) highlighted that while AI has advanced significantly, there are still instances where chatbots produce false and insensitive replies. Additionally, ethical considerations play a significant role in integrating AI in education.

Literature Review

AI chatbots as virtual conversational partners (VCPs) in second language learning

The growing dominance of artificial intelligence (AI) has been witnessed in recent years, especially in second-language education. Several studies have explored its potential implications in teaching English. In a study by Dasborough (2023), regarding its human-like responses, ChatGPT is optimized explicitly for conversational purposes by leveraging its extensive knowledge base to generate verbal text, create artistic content, and synthesize data from various sources for analysis. Moreover, ChatGPT can develop authentic and original knowledge without sounding robotic or artificial. Academic language skills via the use of AI chatbots, such as Quillbot and Chat GPT, have been studied, showing that both students' positive attitudes and language content witnessed notable increases after interacting with these tools over a long period (Truong et al., 2025; Pham, 2024; Su & Tran, 2024)

The notion and framework of chatbots were demonstrated in a study by Gupta et al. (2020), in which they are referred to as a computer program designed to interact with humans through natural, conversational language, and its recent rise in popularity is immensely thanks to significant advancements in artificial intelligence, machine learning, neural networks, and natural language processing. Overall, this study categorizes bots based on the simplicity of their user interface, the algorithms they employ, and the technologies that support their operation, as illustrated in Table 1.

Table 1

Classification of chatbots

Types of chatbots	Functionality
Menu/Button-Based Chatbots	Use buttons and menu options; follow a decision-tree logic. Guide users step-by-step to reach an answer. Easy to use but limited in flexibility and slower in performance.
Keyword Recognition-Based Chatbots	Detect specific keywords in user input and respond based on keyword matches. Efficient for straightforward queries, but may fail with overlapping or ambiguous keywords.
Contextual Chatbots	Use AI and machine learning to understand user intent, analyze sentiment, and learn from previous interactions. Provide personalized, adaptive, and context-aware responses.

According to a study by Duong and Suppasetseree (2024), virtual conversation partners (VCPs) can significantly enhance language learning by providing students with tailored linguistic knowledge and strategies. In this context, Replika, Andy, and Google Assistant were used to record students' self-practice. Multiple studies have shown that VCPs can provide personalized feedback based on students' spoken responses, thereby boosting their willingness to communicate (Wang et al., 2024). Research indicates that VCPs significantly improve students' speaking skills across several linguistic domains, including vocabulary, grammar, pronunciation, and overall fluency. These artificial intelligence (AI) systems assess students' spoken language, identify errors in vocabulary, grammar, and pronunciation, and provide customized suggestions for improvement. Furthermore, AI-driven VCPs provide a more authentic and immersive language acquisition experience. These virtual partners use natural language processing (NLP) to understand and provide authentic responses, allowing students to practice speaking in a setting that closely resembles real-life interactions. Furthermore, VCPs may adjust their remarks based on the learner's feedback, ensuring that the guidance is relevant and customized to the context of the conversation. This adaptability addresses the diverse requirements of learners and accommodates a spectrum of learning preferences (Afkarin & Asmara, 2024)). For students who may feel self-conscious or anxious about conversing in a foreign language, these virtual companions provide a low-pressure, pleasant environment for practicing speaking without fear of judgment or ridicule (Shin & Lee, 2021). These artificial intelligence (AI) technologies provide stimulating and beneficial learning experiences to enhance students' proficiency and confidence in their language abilities. Fathi et al. (2024) assert that the versatility of VCPs allows students to practice at their convenience, enhancing accessibility and accommodating various schedules and preferences. Le et al. (2025) also stressed the importance of automated writing evaluation tools (AWE) in enhancing writing skills for learners.

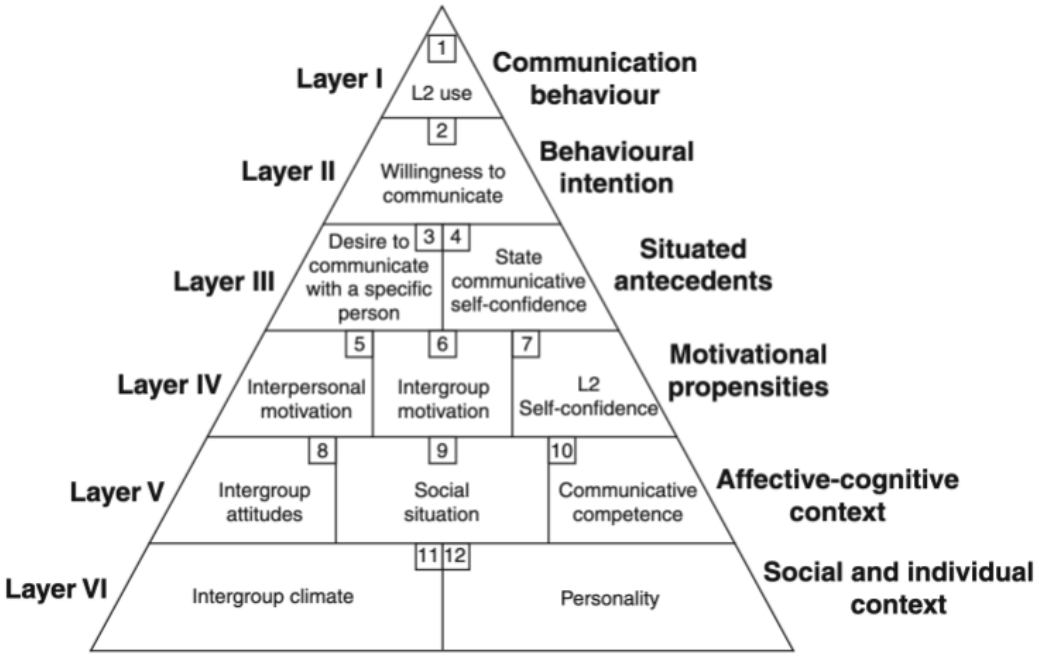
Emotional and behavioral engagement

In English language acquisition, emotional engagement refers to the affective aspect of a student's involvement in learning English as either a second or foreign language. This kind of engagement encompasses the feelings, attitudes, and sentiments experienced by students throughout English language learning, significantly impacting their motivation, perseverance, and overall language acquisition outcomes (Tran et al., 2025; Fathi et al., 2023; Khajavy et al., 2021). Based on these findings, emotional engagement plays a pivotal role in English language acquisition, as it influences students' connection to the language, their commitment to its use,

and their willingness to participate in classroom activities. According to Muñoz et al. (2023), learners with a considerable degree of emotional engagement would convey a greater enthusiasm for learning, demonstrate increased willingness to communicate in English, and show enhanced resilience when confronted with challenges. Conversely, negative emotions such as annoyance and anxiety may impede English proficiency (Tai & Chen, 2024; Yuan, 2024). Language anxiety can lead to diminished participation, avoidance of speaking opportunities, and potentially, language attrition if left unaddressed (Wang & Xue, 2024). Furthermore, emotional engagement is a fluid notion influenced by several factors, including individual characteristics, cultural context, and prior language learning experiences (Fathi et al., 2023; Liu & Wang, 2024). Understanding the intricacies of emotional engagement in English teaching will empower educators to devise more effective strategies to enhance the language learning experience and support the emotional well-being of their students (Nguyen et al., 2025).

Figure 1.

Multi-layered model of variables influencing WTC (MacIntyre, Clément, Dörnyei & Noels, 1998, p. 547)



Dörnyei (2005) highlighted the increased engagement, speaking abilities, and enthusiasm of students for WTC, particularly in light of personalized learning tools. Learners may receive immediate feedback and progressively enhance their communication skills as they interact with artificial intelligence. This increased familiarity may result in an increase in one's desire to participate in real-world scenarios, which would improve their overall language proficiency. WTC also serves as a metric for language usage and influences the effectiveness of language acquisition. As a result, students who are more enthusiastic about speaking are more likely to participate in speaking exercises, which provide them with exposure to and practice in the language (Yashima, 2002). This consistent practice may lead to enhanced fluency, increased confidence in language usage, and stronger linguistic abilities. Conversely, the absence of WTC may result in a decrease in general language proficiency and fewer opportunities for language exposure.

Effects of Generative AI on students' emotional engagement

Numerous studies have gained insights into the impact of Generative AI on learners' emotions in foreign language classrooms, which can include both positive and negative behaviors when communicating with AI chatbots. In a quantitative research study by Khasawneh (2023), 597 undergraduates were selected and studied using a range of AI-powered chatbots, including Linguabot, PoliglotPa, LingvoChat, Speaking AI, and ChatLingo. The concluded results indicated that most students expressed positive opinions of using the applications in their learning or in the future, thanks to the ease of use. They asserted that these tools were their primary aids in enhancing their communication skills in English. In an experimental study by Munoz et al. (2023), 166 students used Chat GPT during the period and scored higher on listening points ($F(2,3) = 4.59$, $p < 0.01$) and intrinsic motivations in learning English ($F(2,3) = 8.047$, $p < 0.001$), analyzed by ANOVA. This indicates an improvement in language learning and students' interests. Moybeka et al. (2023) and Yang et al. (2022) also concluded that AI could enhance students' enthusiasm, self-study and personal learning. In addition, Ellie, an AI chatbot, was researched by Yang et al. (2022), demonstrating that this tool can facilitate conversational practice and engagement. Ni and Cheung (2023) found that intelligent tutoring systems adapt to individual learning styles, enhancing engagement and language acquisition. The rise of AI in English as a Foreign Language (EFL) contexts has led to more studies emphasizing the positive impact of ChatGPT, a leading AI tool, on language skill improvement, learning motivation, and reducing learning anxiety (Jumriah et al., 2022; Tlili et al., 2023). Additionally, ChatGPT has been shown to improve writing efficacy and speaking skills by simulating conversations and providing real-time feedback (Zhou, 2023). However, notwithstanding the apparent benefits of AI tools on emotional engagement, foreign language anxiety (FLA) has also sparked some debate on the prospects of bots minimizing the stress and lack of confidence during direct conversations with programmed tools. El Shazly (2021) observed and investigated the effects of enhanced AI chatbots in classroom activities during pre- and post-intervention in a quasi-experimental study. Surprisingly, 48 participants had witnessed negligible changes in anxiety, confirming the limited role in education, even though a degree of positive attitudes had been noticed towards using AI tools in learning. Pham et al. (2024) concluded that students' engagement levels regarding emotional, behavioral, and cognitive factors would be greatly boosted via the adoption of AI tools in vocabulary retention.

Tai and Chen (2024), Yuan (2024), and Fathi (2024) looked at how Google Assistant affected the willingness to communicate (WTC) of 112 eighth-grade English as a Foreign Language (EFL) students. Engaging Google Assistant with these technologies significantly raised students' willingness to communicate (WTC), communication confidence, and reduced speaking fear, the research revealed. The interactive character of the activities—games and chatbot interactions—which enhanced a more easily accessible learning environment, was appreciated by participants. This led to increased interest and participation, providing evidence of the effectiveness of IPAs in enhancing language learning access. Yuan (2024) examined how well chatbots worked for 74 primary school pupils. Speaking English ability and communication preparedness were gauged using pre- and post-assessments. The experimental group employing chatbots showed significant gains in speaking skills and willingness to communicate when compared to the control group using traditional training methodologies.

Using the Andy English Chatbot, another Fathi (2024) research looked at how AI-mediated speaking activities impacted EFL students' speaking abilities. Sixty-five applicants in total were divided into two groups: artificial intelligence contacts and face-to-face encounters. As for the impacts on speaking fluency, coherence, vocabulary, grammatical accuracy, and pronunciation, the results showed that AI-mediated activities boosted the level of WTC. These results reflect

a considerable inclination for the use of artificial intelligence and chatbots in EFL courses, depending on the above-mentioned WTC-related aspects. This may provide interesting low-stress settings that raise students' overall language ability and communication willingness (WTC).

Concerns regarding the implementation of AI chatbots in language learning

The integration of AI in language education raises several concerns, particularly regarding the quality and authenticity of AI-generated content. Table 2 is a summary of the shortcomings of AI in the world of teaching and learning.

Table 2

Concerns related to AI tools

Key themes	Concerns	Authors
Content authenticity & quality	AI can bring out relevant content, but it often lacks emotional responses, subjective insight, and may include inappropriate or falsified citations.	Auernhammer (2020); Kumar et al. (2023); Chan & Tsi (2023); Chan & Hu (2023)
Skill development limitations	AI struggles to support creativity, critical thinking, and other complex cognitive processes.	Chan & Tsi (2023)
Academic integrity	Students may skip meaningful learning experiences, plagiarize, or violate intellectual property using AI-generated content.	Sok & Heng (2023); Kumar et al. (2023); Pang et al. (2024)
Misinformation & ethical issues	AI may spread misinformation, violate privacy, or contribute to sensitive content.	Huang et al. (2023)
Assessment challenges	Educators may struggle to distinguish between human- and AI-generated responses, leading to biased or inaccurate evaluations.	Sok & Heng (2023)
Emotional impact	Deceptive AI content may reduce students' confidence.	Shiri (2023)

According to Table 2, research by Shiri (2023) and Chan & Tsi (2023) has indicated that despite the efficacy in boosting learning, AI chatbots have been reported to lack specific human aspects in terms of creativity, emotions, and critical thinking ability. They were, in fact, a mimicry of input prompts, which may hinder meaningful experiences or even violate intellectual content (Sok & Heng, 2023; Kumar et al., 2023; Pang et al., 2024). Another issue faced by learners involves inaccurate and biased information (Sok & Heng, 2023), in which cautions ought to be made to ensure the reliability and validity of output.

Call Annie

The extension of Chat GPT into a more human-like text-based chatbot has sparked numerous insights. This app utilizes GPT-3.5 for its conversational abilities and offers a lifelike chat experience through natural language processing and machine learning. According to Sharma (2023), instead of generating instant messages based on provided prompts, the application "Call Annie", developed by Animato, was programmed from the "conversational capabilities" of Generative AI to produce audio voice texts, making the textual dialogues resemble two-way communication between two interlocutors. It is accessible on both iOS and Android devices, as well as via the web at https://callannie.ai. In light of the verbal sentences spoken, the tool can be accessed via the click of both video and audio phone calls. The default character is Annie, a 30-year-old woman, but users can choose from various other characters with distinct

personalities and speaking styles. Video chats can be started by selecting the 'Call [Character Name]' button, and the AI will begin the conversation, continuing if the user engages. The app also provides 'conversation ideas' and a 'learn' mode for practicing specific scenarios, though most are behind a paywall. To access more content, users can subscribe to Call Annie Super for $11.99 per month or $79.99 annually, in addition to the free charge during use (Wan & Moorhouse, 2024).

The introduction of Call Annie into the teaching and learning of English

Amidst the breakthrough of Generative AI in education, Call Annie has been rendered a potential tool for English learners (Wan & Moorhouse, 2024). Nguyen et al (2025) had claimed that despite students' positive attitudes towards the use of Call Annie, there had been significant concerns that educators should be cautious of. Thanks to the diverse scenarios and lifelike characters (Sharma, 2023), language instructors can explore ways to incorporate Annie into the classrooms. Moreover, to enable the issue of self-directed learning in language acquisition, teachers can use a scaffolded approach to encourage students to use the tool in various situations (Younas et al., 2025). For instance, instructors can assist students in creating a personalized learning plan with Call Annie by helping them establish specific language learning objectives that align with course learning outcomes and providing suitable prompts during interactions. Wan & Moorhouse (2024) also emphasized that in the classroom, teachers can have students engage in role-playing exercises with 'Annie' related to the lesson's topic, which might offer students a low-pressure environment to make mistakes, learn, and receive feedback before interacting with classmates. Teachers can assign speaking exercises outside of class as homework to reinforce the skills learned.

Research gaps

Despite the attention given to the rise of AI chatbots, research aimed at exploring the efficacy of “Call Annie" has remained elusive, and there is still a limited study that examined the relationship between students' emotional engagement and the use of an AI chatbot in English classrooms. In Vietnamese contexts, a number of students learn English in university contexts without the constant exposure to authentic English in their daily lives. Unfortunately, many learners might experience concerns and anxiety because they lack confidence when communicating in English. Thus, this study aimed to investigate the impact of the Call Annie chatbot on students' emotional engagement when practicing speaking skills. This requires an extended investigation into students' perception and emotional engagement with the direct application. Further research is, in turn, needed to assess the true effectiveness of 'Call Annie' and other GAI chatbots, which could shed light on the future potentialities for language educators.

Research Questions

This study aims to answer the following two research questions:

(1) How does using Call Annie as a virtual conversational partner affect English-majored students' emotional engagement in speaking lessons?

(2) How does students' behavioural engagement or WTC change after using Call Annie in personalized conversations?

Methods

Pedagogical Setting & Participants

As shown in **Table 3**, the population of this research was 95 English-majored students from 2 speaking classes at the Faculty of Foreign Language Department at a university. The sample consisted of 95 students, of whom 18 students were excluded due to their inconsistent survey scores. The majority of students are sophomores, whereas only a negligible number are in their third or fourth years.

Table 3

Demographic Data of Study Participants

Attributes	Frequency	Percentage
Gender		
Female	57	74,03%
Male	20	25,97%
Total	77	100%
Year of study		
Sophomore	73	94,81%
Junior	3	3,9%
Senior	1	1,3%

In general, 77 English majors, whose English levels are mostly B1, came from 2 speaking classes at the Faculty of Foreign Language department of a university. They recorded regular conversations with “Call Annie” via Zoom in a three-week homework assignment project. They were asked to complete 6 video recordings. Among them were 57 female and 20 male students, making up 74,03% and 25,97%, respectively, of the overall population.

William (2007) highlighted the importance of using survey questions in research, as they can provide validated and reliable data from participants who are asked similar questions. The data was collected through a Google form of students' emotional engagement, pre- and post-questionnaire, and open-ended questions.

Design of the Study

This research employed a mixed methods approach to investigate the impact of using Call Annie in speaking homework assignments on students' emotional engagement levels, focusing on three key factors: confidence, anxiety, and willingness to communicate, which are crucial for motivating students to speak English as a foreign language at the university level. Students' engagement could be collected based on their responses to questionnaires adopted from Chan & Hu (2023) that consisted of 17 questions, including both closed and open-ended questions. Data from pre- and post-intervention using Call Annie were analyzed using an appropriate test in the IBM statistical package and service solution (SPSS 22) to assess the application's significance as a treatment in this research, which could noticeably enhance students' engagement level in conversing in English.

Research instrument

The quantitative data of this study is aimed at exploring the changes in behaviors before and after the use of Call Annie for homework completion. Pre and post-survey results are used to measure the effect of the application on students' emotional aspects. The hypothesis testing is listed out as follows:

1/ **One/ Null Hypothesis (H0):** There is no significant improvement in students' emotional engagement, perceived concerns, and willingness to communicate after adopting Call Annie for speaking practice.

2/ **Two/ Alternative Hypothesis (H1):** There is a significant improvement in students' emotional engagement, perceived concerns, and willingness to communicate after adopting Call Annie for speaking practice.

To illustrate the significant data, the p-value of the t-test is under 0.05, meaning that the null hypothesis is rejected and the tool has a great effect on students' engagement levels.

Data collection & analysis

Regarding emotional engagement, concerns, and willingness to use the app to communicate in the future, this research employed and adapted survey questions developed by Chan & Hu (2023). This instrument is represented by three scales, measuring students' perceptions of their emotions, concerns, and future tendencies to continue using Call Annie for regular speaking practice. The scale consisted of 21 items, using a five-point Likert scale, ranging from "strongly disagree" to "agree strongly".

Table 4

Reliability of questionnaire

Cronbach's alpha	Number of items
.965	17

The reliability of the instrument with 17 items using Cronbach's Alpha was 0.965, as shown in **Table 4**. The reliability of each subscale is also presented in Table 6, at 0.985, 0.664, and 0.938, respectively.

Table 5

Reliability of each subscale

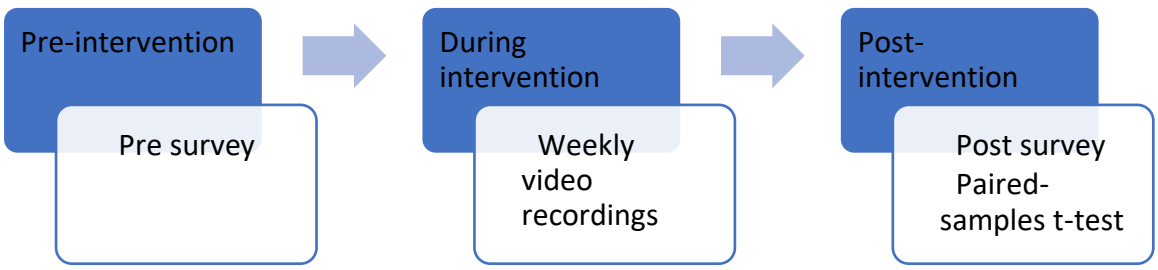
Scale	Number of items	Reliability
Emotional engagement	9	.958
Perceived concerns	3	.664
Willingness to use	5	.938

Data collection technique

The collection of evidence backing up for this study is implemented in the following phases (see figure 2)

Figure 2.

Procedure for the data collection period



Details about the sequence of gathering data from participants are explained below in Table 6.

Table 6

Procedure for the data collection period

Phase 1: Questionnaires for pre-intervention	During week 1, questionnaires for the pre-intervention were distributed to students in two speaking classes to gather information about their emotional engagement level, perception of the use of Gen AI in homework assignments, and concerns. Instructions on how to use the app and guidelines for role-play tasks were also introduced for students' participation. Participants were expected to answer multiple-choice questions and open-ended questions on a Google form: • Questions 1 and 2 collect the participants' biographical information, including their name and year of university. This information would help the researchers better understand the students' backgrounds and thereby decide on suitable samples for the research. • Questions 3 to 21 explored various aspects of students' emotions, concerns, and their willingness to use the app in the future to speak English, specifically in the context of implementing Call Annie.
Phase 2: Video recordings for weekly homework assignments	After collecting quantitative data via questionnaires, students were assigned to conduct weekly projects, performing different roles in virtual conversations with "Call Annie" in at least six videos (with the option to include additional videos for individual participation). The contexts ranged from a job interview to a conversation between a tutor and a struggling student learning English, meaning that students had a choice to opt for language items they needed to complete the tasks. Below is the extract from the assigned roles and suggested questions for students' interactions with the tool. **Then, ask Wh- questions in detail and reply to Annie's questions if any. You should be polite and interactive as this should be a two-way conversation.** **Content** **Week 1**: *Video 1:* Role-play a job interview applying for teachers/teaching assistants alone or in groups of 2, 3 or 4 students with Annie, asking questions given by the teacher (Students = managers, Annie = candidates) *Video 2:* Role-play an interview with a native speaker about Vietnam, asking Annie the following questions: 1/ What do you think about Vietnam? 2/ Have you ever thought of staying in Vietnam forever? 3/ Could you suggest some ways I can improve my English? 4/ Tell Annie about your day and ask if she can comfort you. **Week 2:** *Video 3*: Role-play a conversation between a personal tutor (Annie) and a student (You) in class. You are having problems with your pronunciation and you cannot pronounce a few words. Ask Annie how you could improve your pronunciation (Suggested words: wreak havoc, broaden my horizon, life expectancy, indicators of a country's prosperity,...) *Video 4:* Ask Annie about the pros and cons of tourism and take notes. **Week 3:** *Video 5,6,...:* Choose a random topic and create a conversation with Annie. You can upload as many videos as you like. (Tourism, vacation plan, mindfulness, healthy lifestyle,...) Subsequently, participants had their speeches recorded via Zoom and submitted the link on a Google sheet, delivered by the lecturer.

Phase 3: Questionnaires for post-intervention	After week 3, questionnaires for post-intervention were completed by the engaged students in order to track the changes in their emotional performance. The same set of questions was utilized and compared later, using Paired-Samples T-Test.

Data analysis technique

The data from questions 1 and 2 was synthesized using Excel and tables to present students' background information. Evidence of students' answers from questions 3 to 19 during pre- and post-intervention was analyzed via the analysis of Paired-Samples T-Test, SPSS 22, to assess the changes in students' perception over the investigated period. The answers collected from the open-ended questions 21 and 22 were first summarized and then coded following the thematic analysis guidelines. A comparison between quantitative and qualitative data was reflected to explore the results (Cress-well & Plano Clark, 2011)

Results/Findings

The descriptive results of participants' performance on pre- and post-intervention towards using Call Annie are shown in Table 7. An initial impression from the chart is that the mean scores range from 3.429 to 3.573 in the pretest. At the same time, in the post-results, the level is noticeably higher, with an equivalent point of more than 4.0 in the level of emotional engagement and willingness to use the app, and 3.696 distributed for the rate of concerns.

Table 7

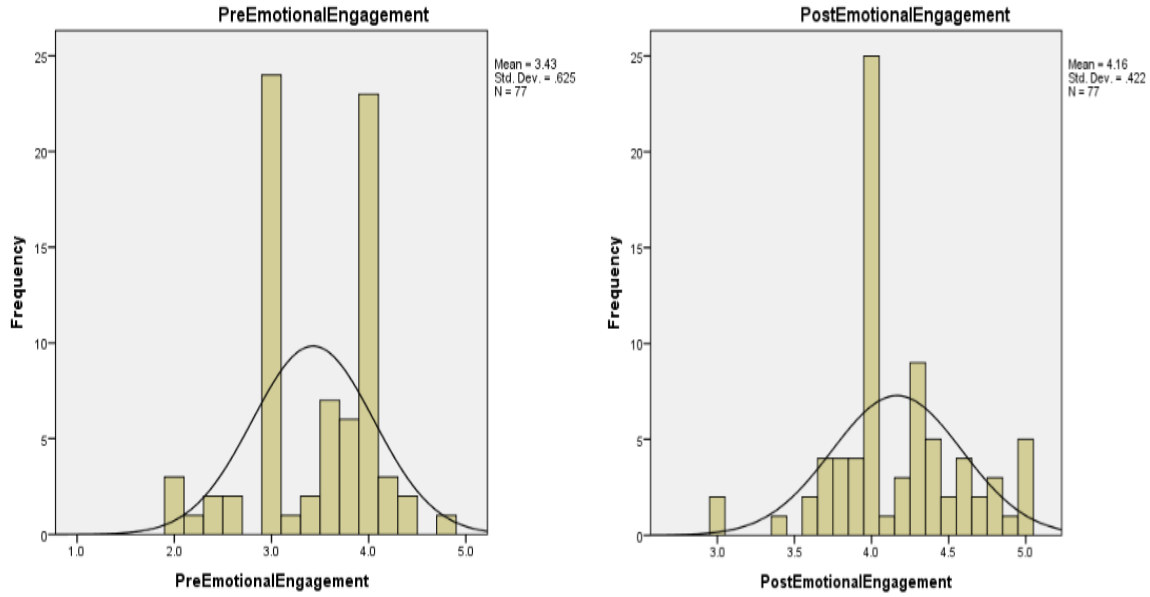
Descriptive analysis of pre and post-intervention

Domains	N	Minimum	Maximum	Mean	SD	Skewness	Kurtosis
Pre-Emotional engagement	77	2.0	4.7	3.429	.6247	-.463	-.433
Pre-Concerns	77	1.0	5.0	3.430	.7749	-.260	.411
Pre-Willingness to use	77	2.0	4.8	3.573	.6303	-.313	-.481
Post-Emotional engagement	77	3.0	5.0	4.165	.4220	.022	.508
Post-Concerns	77	1.0	5.0	3.696	.6375	-.997	3.003
Post-Willingness to use	77	3.0	5.0	4.171	.4261	-.172	.212

According to Bryne (2010), the data distribution is rendered normal and fulfilled only when skewness and kurtosis coefficients are within the range of ±2 and ±7, respectively. In Table 7, the skewness values of most scales are negative, except for the rate of post-emotional engagement. In contrast, in the kurtosis coefficients, only pre-emotional engagement and willingness to use are negative. In general, the numbers for skewness and kurtosis in this survey are all acceptable in the range of normal distribution. As for the frequency of the data of 3 domains, the histograms in Figures 3, 4, and 5 illustrate a considerable improvement in their enjoyment and confidence levels after using Call Annie for their speaking practice.

Figure 3

Frequency statistics for emotional engagement level

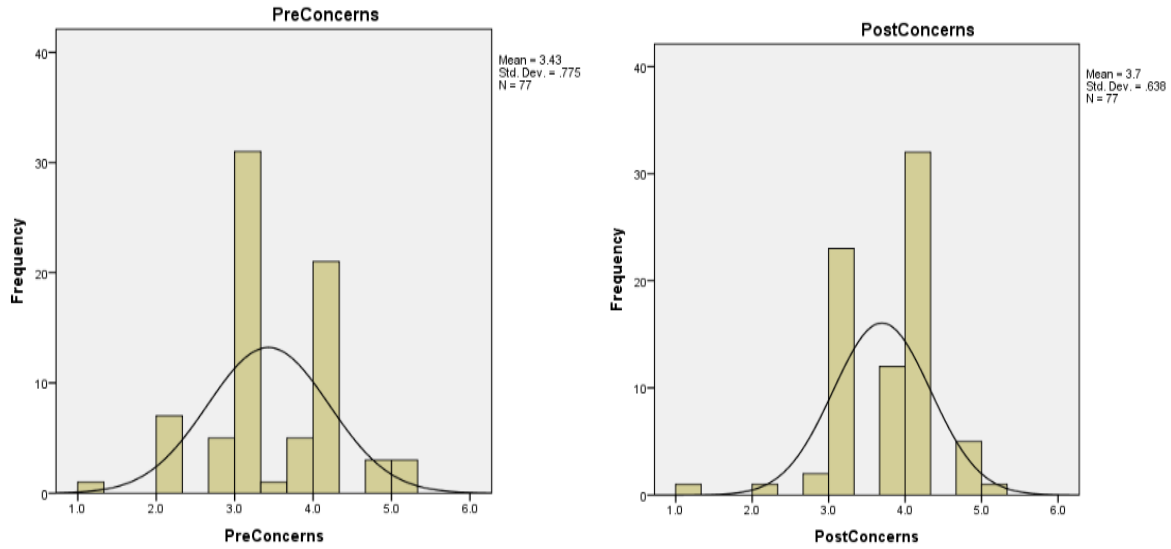


In **Figure 3**, pre-test scores (N = 77, M = 3.43) show a relatively broad distribution, with most students scoring between 3 and 4 points. Meanwhile, the post-test scores (M = 4.16) depict higher average positive emotions and more consistent points.

Regarding ethical concerns associated with Call Annie as an AI chatbot in **Figure 4**, the data shows that after an intervention, the mean score increased from 3.43 (pre-intervention) to 3.7 (post-intervention).

Figure 4

Frequency statistics for the level of ethical concerns towards Call Annie as an AI chatbot



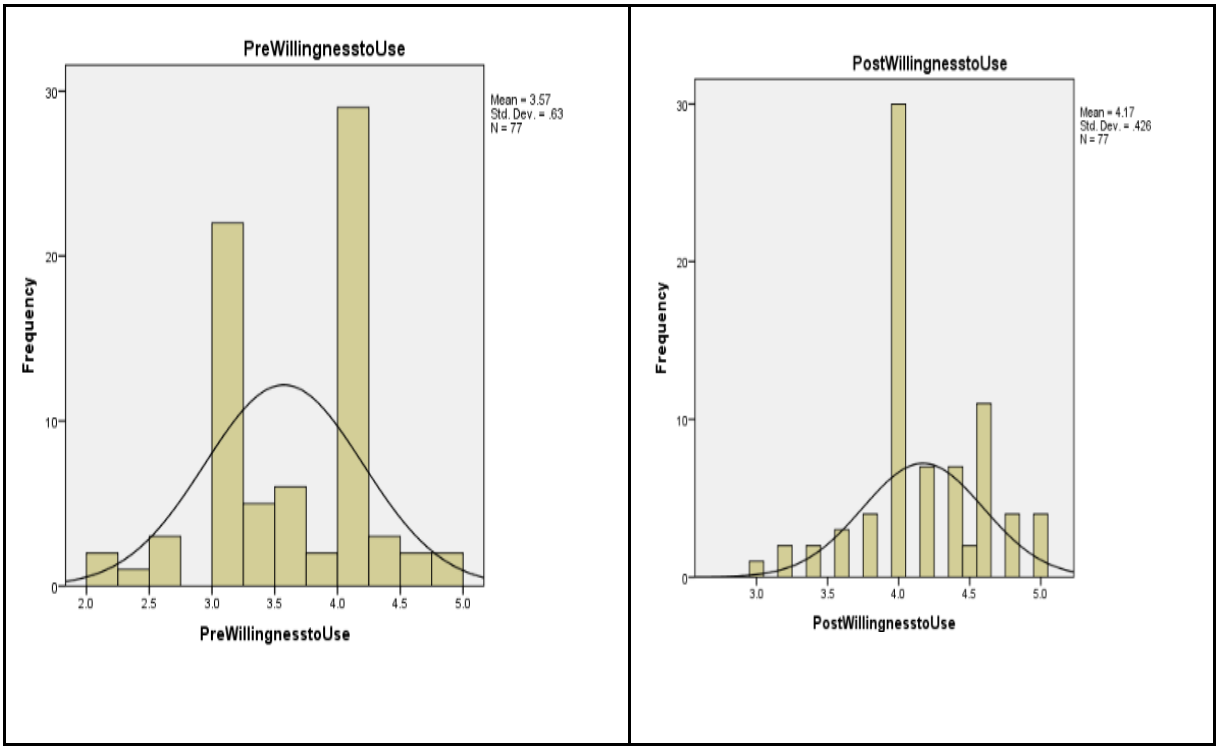
The evidence in Figure 4 suggests that after 3 weeks, students perceived higher ethical concerns about the chatbot as an increase in the mean score indicates a shift in opinion or awareness,

possibly implying that the intervention (such as bias, accuracy, or lack of human features) made users more aware or cautious about ethical issues related to the AI's usage.

Figure 5 highlights the level of willingness to use Call Annie in the future; the mean score increased from 3.57 (pre-intervention) to 4.17 (post-intervention), which might indicate that participants became more willing to use Call Annie as an AI chatbot after the 3-week trial.

Figure 5

Frequency statistics for the level of willingness to use the tool in the future



In **Figure 5**, a higher mean score suggests that the virtual interactions with Call Annie positively influenced users' attitudes toward the chatbot, making them more likely to use it in the future.

Table 8

Comparison of Mean scores of pre-test and post-test scores

Domains	N	Correlatio	Sig
Pre & Post Emotional engageme	77	.422	.000
Pre & Post Concerns	77	.233	.041
Pre- and Post-Willness to use	77	.516	.000

*Significant at α=0.05

In Table 8, a paired samples t-test was conducted to assess the impact of the invention on emotional aspects, comparing pre- and post-use survey results after personal adoption of Call Annie. These paired sample correlations depict a moderate and positive correlation between the pre- and post-scores for all domains (with the correlation values being 0.422, 0.233, and 0.516 for emotions, concerns, and willingness to use. Significant values are over 0.001, underscoring the consistency in the individual scores.

Table 9

Comparison of Mean scores of pre-test and post-test employing Sample Test - Paired Differences

	Mean	SD	SD Error Mean	95% Confidence Interval of Difference		Sig. (2-tailed)
				Lower	Upper	
Pre & Post Emotional engagement	-.7364	.5880	.0670	-.8698	-.6029	.000
Pre & Post Concerns	-.2662	.8812	.1004	-.4662	-.0662	.010
Pre- and Post-Willness to use	-.5987	.5490	.0626	-.7283	-4.4741	.000

*Significant at α=0.05

From an emotional standpoint, **Table 9** illustrates a statistically significant mean difference between pre-and post-survey results (M = -.7364, SD = .5880, 95% CI [-.8698, -.6029], p = .000. These findings provide robust evidence of a significant improvement in positive emotion levels following the intervention. The figures for concerns and readiness to use Call Annie also witnessed noticeable rises, with p-values being below 0.05, asserting the significance of the changes.

Pre-intervention

Before the intervention, participants expressed mixed reviews of the Call Annie app, indicating both joy and doubt. About 65% of those who participated said the tool was "very useful," "helpful," or "convenient," and they often talked about how it helped them improve their language, speech, speaking confidence, and listening skills. Several students, especially those who were shy or did not want to speak up in class, said that practicing made them feel more at ease.

On the other hand, about 35% of comments were negative, mostly about how real AI conversations seem. In addition, while there were reportedly a number of technical problems, the chatbot could not always recognize speech or only react in a limited number of ways. Lack of emotional depth and interactivity was another point made by the students.

Post-intervention

On the whole, there had been noticeable changes after the period of interacting with Call Annie. About 75% of students reported that the improved experience using the tool to practice speaking was beneficial. Many said they felt more comfortable speaking in public and liked that they could practice whenever they wanted without having to find a partner. The app's low-pressure setting continued to be useful for students who are shy or quiet. However, about 25% of users still reported ongoing problems associated with speed, robotic behaviors, involving inaccurate answers, and a lack of emotional support.

Nonetheless, many students supported the idea of incorporating such tools into language classrooms. This was shown in comments such as “teachers should use this in speaking classes” and “students should be encouraged to try it.” Some participants suggested that the app could also be beneficial for younger learners. Others recommended continued teacher involvement—offering feedback, support, and guidance to ensure that the tool supports rather than replaces the role of teachers.

Discussion

With reference to the p-value from t-test analysis, students' interests in learning had been greatly boosted thanks to the use of AI tools, which aligns with the findings from Khasawneh (2023) and Muñoz et al. (2023). This may be due to the relaxed, low-pressure environment provided by the app, where students felt more willing to take risks without the fear of being judged. In essence, although an overwhelming majority of students would like to continue using the app in the future, they also reported a greater level of caution in this context. Such caution could stem from their awareness of certain limitations, such as the lack of emotional depth or concerns about over-reliance on non-human interactions. This highlight would be useful for educators when adapting their lessons with the incorporation of Call Annie. It is therefore important that teachers take into account not only the benefits of AI but also the possible hesitations students may have when navigating these new learning modes. Considering the ethical dilemmas, there has been a growing interest in AI technology in speaking English; human-to-machine relationships are also a pivotal phenomenon that teachers should take into account. These interactions, although convenient, may still feel unnatural or emotionally flat for some learners. The observation that users showed a greater willingness to engage after using Call Annie is consistent with the findings of Tai & Chen (2024) and Yuan (2024), which suggest that AI-powered applications may enhance individuals' confidence in their interactions with others. This can be explained by the safe, non-judgmental space the app offers, allowing users to rehearse conversations at their own pace. In addition, the results of this study also suggest that students have the opportunity to practice their speaking skills in a supportive environment with robots, where they are free from any judgment. This, in turn, may lead to more confidence when speaking with real people. They are certainly more adept at engaging with others face-to-face. In addition, the level of WTC rose to 4.17, which confirms the claim from a study by Fathi (2024) and reports on students' increased readiness to participate in conversations. Moreover, based on research completed by Chan and Hu (2023), students might exhibit a higher extent of WTC, whereas their concerns related to authenticity and emotional challenges from the conversational agent were inevitable, as asserted by the results of this study. These concerns should not be overlooked, as they reflect a need for emotional connection and natural responses, which current AI-powered tools may not fully provide. Anxiety levels, likewise, saw notable increases following the intervention. This might be a result of heightened self-awareness or unmet expectations when students realized that AI responses lacked emotional elements. The students experienced factors related to accuracy, bias, and the absence of human aspects, in addition to enthusiasm, aspirations, and educational benefits, which aligns with Kumar's (2023) research findings.

Examining answers from the open-ended questions made before and after the experimental period allows us to better understand how individuals view the role of AI chatbots, such as Call Annie. AI robots create a situation conducive to language practice, benefiting a wide range of individuals, particularly those who may be more reserved. This suggests that learners who typically hesitate to speak up in traditional classroom settings may find a sense of ease when interacting with a non-human partner, as it reduces social anxiety and fear of being judged. By the same token, the students encountered a lack of emotional support and clumsy responses from Call Annie, which could partially hinder the engagement between human and AI agents. This highlights the fact that while the AI offers practical assistance, it may still fall short in building a meaningful connection, which is often crucial in sustaining long-term motivation. This resembles the results from a study by Chan and Tsi (2023), indicating that emotional resonance remains a challenge in AI-mediated learning. While some individuals believe that AI chatbots can provide help in language learning, they argue that it is unable to offset two essential

human abilities: critical thinking and creativity. This indicates the gap between mechanical interactions and human contact. Finally, a proportion of students are willing to utilize AI technologies as personal assistants in the future for their study, instead of replacing human teachers, as asserted by Ni and Cheung (2023). This preference reflects a balanced mindset among learners, who seem to welcome AI as a supportive tool while still valuing the irreplaceable role of human educators in guiding, inspiring, and personalizing their learning experiences.

Conclusion

Overall, the aim of this study is to gauge the levels of emotional engagement based on three factors regarding emotional and behavioral attitudes, together with the students' perception of concerns towards the use of Call Annie as a virtual conversational partner for homework assignments. In short, the results from the sample-paired t-test have depicted the increased degree of affective elements, which were crucial for personal language development. Students reported that they felt more at ease speaking English, which is a critical component of acquiring any language. This was a significant finding. The study also noted that students were more inclined to communicate, a critical component of language acquisition that many students encounter difficulties with. This willingness to engage with an AI companion, despite the peculiar setting, suggests that students were more motivated to enhance their speaking abilities and felt more at ease. Notwithstanding their ongoing concerns related to privacy and validity of information, students are expressing positive hopes for the integration of AI tools into the curriculum, as it can compensate for unwilling or shy learners.

Limitations of the Study

In spite of perceived advantages, certain cautions should be acknowledged. Although most learners indicated enthusiasm for the application, their emotional commitment could not be rendered the sole determinant of improved proficiency in English. Therefore, there is a need for an experimental approach to evaluate learners' proficiency regarding Call Annie's adoption in enhancing speaking skills.

Another area for development is the extremely limited sample size of 85 students, which may not accurately generalize to the broader population of English language learners. It is recommended that a larger sample size would enable the development of more reliable findings on the use of AI technology in language learning. To examine the long-term impact of the tool, the 3-week duration of future research should be extended to promote the effectiveness, along with an experimental design in measuring the level of linguistic mastery, instead of emotional aspects.

Suggestions for further studies

Longitudinal and experimental research designs would be of importance to evaluate students' levels before and after the intervention of AI-powered tools. What is more, with the aim of striking a balance between students' demands for personalized learning strategies and ethical issues associated with human content, educators would be expected to incorporate training for learners to maximize their learning outcomes, especially for introverted participants who have a limited level of English. A high extent of WTC could synonymously boost the engagement of students in homework assignments, thereby allowing them to interact in English with the AI chatbot in the future.

References

Afkarin, M. Y., & Asmara, C. H. (2024). Investigating the Implementation of ChatGPT in English Language Education: Effects on Student Motivation and Performance Levels. *Journey: Journal of English Language and Pedagogy*, *7*(1), 57–66. https://doi.org/10.33503/journey.v7i1.865

Auernhammer, J. (2020). Human-centered AI: The role of Human-centered Design Research in the development of AI. https://doi.org/10.21606/drs.2020.282

Bao, D. (2019). The place of silence in second language acquisition. *English Language Teaching and Research Journal* (ELTAR-J), *1*(1), 26-42.

Byrne, B. M. (2010). Basic concepts, applications, and programming. *Structural equation modeling with AMOS*, *396*, 227.

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1), 43. https://doi.org/10.1186/s41239-023-00411-8

Chan, C. K. Y., & Tsi, L. H. (2023). The AI revolution in education: will AI replace or assist teachers in higher education? https://doi.org/10.48550/arXiv.2305.01185

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Dasborough, M. T. (2023). Awe-inspiring advancements in AI: The impact of ChatGPT on the field of Organizational Behavior. *Journal of Organizational Behavior (John Wiley & Sons, Inc.)*, *44*(2). https://doi.org/10.1002/job.2695

Duong, T., & Suppasetseree, S. (2024). The effects of an artificial intelligence voice chatbot on improving Vietnamese undergraduate students' English speaking skills. *International Journal of Learning, Teaching and Educational Research*, *23*(3), 293-321. https://doi.org/10.26803/ijlter.23.3.15

Dornyei, L. (2005). The psychology of the language learners: Mahwah. *New Jorzy: Laurence Ulbanm Associates*. https://tesl-ej.org/wordpress/issues/volume10/ej37/ej37r7/?wscr=

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, *38*(3), e12667. https://doi.org/10.1111/exsy.12667

Fathi, J., Derakhshan, A., & Ebrahimi Galedar, D. (2023). An investigation of the relationship between teacher self-efficacy and teacher academic optimism among Iranian EFL teachers: A structural equation modelling approach. *Language and Translation Studies* (LTS), *55*(4), 33-63. https://doi.org/10.22067/jlts.2021.71751.1060

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, *121*, 103254. https://doi.org/10.1016/j.system.2024.103254

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer assisted language learning*, *27*(1), 70-105. https://doi.org/10.1080/09588221.2012.700315

Gupta, A., Hathwar, D., & Vijayakumar, A. (2020). Introduction to AI chatbots. *International Journal of Engineering Research and Technology*, *9*(7), 255-258.

Huang, L. (2023). Ethics of artificial intelligence in education: Student privacy and data protection. *Science Insights Education Frontiers*, *16*(2), 2577-2587. https://doi.org/10.15354/sief.23.re202

Jumriah, J., Susilawati, E. S., Supriatna, E., Smas, M. H., & Arini, I. (2024). Analysis of the Use of GPT Chat to Improve Student Performance. *EDUKASIA: Jurnal Pendidikan Dan Pembelajaran*, *5*(1), 127-136. https://doi.org/10.62775/edukasia.v5i1.725

Kang, S., & Sung, M. C. (2024). EFL students' self-directed learning of conversation skills with AI chatbots. *Language Learning & Technology*, *28*(1), 1-19 (ISSN: 1094-3501). https://hdl.handle.net/10125/73600

Khasawneh, M. A. S. (2023). Integrating Ai-Based Virtual Conversation Partners in Enhancing Speaking Skills In Foreign Languages: Insights From University Students. *Journal of Southwest Jiaotong University*, *58*(5), 575-587. http://jsju.org/index.php/journal/article/viewFile/1861/1850

Khajavy, G. H., MacIntyre, P. D., & Hariri, J. (2021). A closer look at grit and language mindset as predictors of foreign language achievement. *Studies in Second Language Acquisition*, *43*(2), 379-402. https://doi.org/10.1017/S0272263120000480

Kumar, M. V., Ramesh, G. P., Pareek, P. K., Deepak, H. A., & Babu, J. A. (2023, June). Robotic Attendance Scheme in the Classroom Using Artificial Intelligence and Internet of Things. In *2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC)* (pp. 1-6). IEEE. https://dialnet.unirioja.es/servlet/articulo?codigo=9869857

Le, T., Nguyen, X., Dang, H., Pham, T., & Nguyen, L. (2025). Automated Writing Evaluation (AWE) in Vietnamese EFL Classrooms: Enhancing skills and understanding Student perception. *Journal of Science and Technology - Industrial University Ho Chi Minh City*, 73 (ISSN: 2525-2267). https://jst.iuh.edu.vn/index.php/jst-iuh/article/view/5095

Liu, W., & Wang, Y. (2024). The effects of using AI tools on critical thinking in English literature classes among EFL learners: An intervention study. *European Journal of Education*, *59*(4), e12804. https://doi.org/10.1111/ejed.12804

Marmolejo-Ramos, F., Workman, T., Walker, C., Lenihan, D., Moulds, S., Correa, J. C., ... & Sonna, B. (2022). AI-powered narrative building for facilitating public participation and engagement. *Discover Artificial Intelligence*, *2*(1), 7.

MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The modern language journal*, *82*(4), 545-562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

Muñoz, S. A. S., Gayoso, G. G., Huambo, A. C., Tapia, R. D. C., Incaluque, J. L., Aguila, O. E. P., ... & Arias-Gonzáles, J. L. (2023). Examining the impacts of ChatGPT on student motivation and engagement. *Social Space*, *23*(1), 1-27.

Ni, A., & Cheung, A. (2023). Understanding secondary students' continuance intention to adopt AI-powered intelligent tutoring system for English learning. *Education and Information Technologies*, *28*(3), 3191-3216. https://doi.org/10.1007/s10639-022-11305-z

Nguyen, L. A. D., Dang, V. M. T., & Le, M. T. (2025). An investigation into Students' Perception of Using Call Annie as a Virtual Conversational Partner. *Proceedings of the AsiaCALL International Conference*, 6, 344-364. ISSN: 2833-6836, ISBN: 979-8-9870112-6-3. DOI: https://doi.org/10.54855/paic.24624

Nguyen, V. L. (2024). The Impact of Online Resources via Personal Devices on English-Majored Students' Learning Practices. *International Journal of Language Instruction*, *3*(4), 110-118. https://doi.org/10.54855/ijli.24346

Pham, T. T., Nguyen , L. A. D., Dang, H. M., & Le, T. T. P. (2024). Exploring Tertiary Vietnamese EFL Students' Engagement in Vocabulary Learning through the Use of an AI Tool. *Proceedings of the AsiaCALL International Conference*, 4, 129–149. https://doi.org/10.54855/paic.23410

Pham, N. Q. N. (2024). The Perspectives of Post-graduates Majoring in English Regarding the Usage of QuillBot to Enhance Paraphrasing Skills. *International Journal of AI in Language Education*, *1*(1), 29-40.https://doi.org/10.54855/ijaile.24113

Pang, S., Nol, E., & Heng, K. (2024). Generative AI as a personal tutor for English language learning: A review of benefits and concerns. https://ssrn.com/abstract=4865398 or http://dx.doi.org/10.2139/ssrn.4865398

Rad, H. S., R. Alipour, and A. Jafarpour. 2023. "Using Artificial Intelligence to Foster Students' Writing Feedback Literacy, Engagement, and Outcome: A Case of Wordtune Application." *Interactive Learning Environments*: 1–21. https://doi.org/10.1080/10494820.2023.2208170.

Sharma, B. (2023). Call Annie: New App Brings ChatGPT To Life With Human-Like Conversations And An Avatar. https://www.indiatimes.com. Retrieved April 11, 2024, from https://www.indiatimes.com/technology/news/call-annie-new-app-brings-chatgpt-to-life-601056.html

Shin, D., Kim, H., Lee, J. H., & Yang, H. (2021). Exploring the use of an artificial intelligence chatbot as second language conversation partners. *Korean journal of English language and linguistics*, *21*(21), 375-391. https://doi.org/10.15738/kjell.21..202104.375

Shiri, A. (2023) ChatGPT and Academic Integrity. Information Matters. https://informationmatters.org/2023/02/chatgpt-and-academic-integrity/

Sok, S., & Heng, K. (2023). ChatGPT for education and research: A review of benefits and risks. *Cambodian Journal of Educational Research*, *3*(1), 110-121.

Su, A. A. T., & Tran, T. H. N. (2024). The Benefits and Weaknesses of Chat GPT on Students' Learning Writing. *International Journal of AI in Language Education*, *1*(1), 20-28.https://doi.org/10.54855/ijaile.24112

Tai, T. Y., & Chen, H. H. J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education*, *220*, 105112. https://doi.org/10.1016/j.compedu.2024.105112

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart learning environments*, *10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Tran , H. N., Le, T. T. N., & Tran, V. B. U. (2025). AI Tools in Learning Academic Writing: Benefits and Challenges for MA Students in the English Language Studies at the Industrial University of Ho Chi Minh City. *International Journal of AI in Language Education*, *2*(1), 74-91. https://doi.org/10.54855/ijaile.25215

Truong, T. A. A., Le, H. K. N., & Nguyen, V. H. Q. (2025). English-Major Master's Students Regarding the Use of ChatGPT in Learning Research Writing at IUH. *International Journal of AI in Language Education*, *2*(1), 92-115. https://doi.org/10.54855/ijaile.25216

Wan, Y., & Moorhouse, B. L. (2024). Using Call Annie as a generative artificial intelligence speaking partner for language learners. *RELC Journal*, 00336882231224813. https://doi.org/10.1177/00336882231224813

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

Wang, Y., & Xue, L. (2024). Using AI-driven chatbots to foster Chinese EFL students' academic engagement: An intervention study. *Computers in Human Behavior*, 159, 108353. https://doi.org/10.1016/j.chb.2024.108353

Xu, X., Dugdale, D. M., Wei, X., & Mi, W. (2023). Leveraging artificial intelligence to predict young learner online learning engagement. *American Journal of Distance Education,* 37(3), 185-198. https://doi.org/10.1080/08923647.2022.2044663

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

Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *The modern language journal*, *86*(1), 54-66. https://doi.org/10.1111/1540-4781.00136

Younas, M., El-Dakhs, D. A. S., & Jiang, Y. (2025). A comprehensive systematic review of AI-driven approaches to self-directed learning. *IEEE Access*. https://ieeexplore.ieee.org/abstract/document/10906568

Yu, H., Liu, Z., & Guo, Y. (2023, April). Application Status, Problems and Future Prospects of Generative ai in education. In 2023 *5th International Conference on Computer Science and Technologies in Education (CSTE)* (pp. 1-7). IEEE. https://doi.org/10.1109/CSTE59648.2023.00065

Yuan, Y. (2024). An empirical study of the efficacy of AI chatbots for English as a foreign language learning in primary education. *Interactive Learning Environments*, *32*(10), 6774-6789. https://doi.org/10.1080/10494820.2023.2282112

Zhou, W. (2023). Chat GPT Integrated with Voice Assistant as Learning Oral Chat-based Constructive Communication to Improve Communicative Competence for EFL learners. https://doi.org/10.48550/arXiv.2311.00718

Biodata

Anh-Duong Nguyen Lam has obtained 10 years of teaching the language to a diverse population of learners in a number of academic contexts (universities, businesses, language centers) until now. The areas of her research interests are using technologies in teaching English, curriculum designs and methods in teaching English skills.

Thien-Phuoc Le is an English lecturer at the Faculty of Foreign Languages at the Industrial University of Ho Chi Minh City, Vietnam. She has more than fifteen years of teaching experience in linguistics and language skills. Her research interests include lifelong learning, student engagement, project-based learning, and technology for sustainable learning.