Belda-Medina and Kokošková
Int J Educ Technol High Educ (2023) 20:62
https://doi.org/10.1186/s41239-023-00432-3

International Journal of Educational Technology in Higher Education

RESEARCH ARTICLE Open Access

Integrating chatbots in education: insights from the Chatbot-Human Interaction Satisfaction Model (CHISM)

Jose Belda-Medina[1*] and Vendula Kokošková[2]

*Correspondence:
jr.belda@ua.es

[1] Department Filología Inglesa, Facultad de Letras, Campus de San Vicente, University of Alicante, 03690 San Vicente, Alicante, Spain
[2] Institute of Applied Language Studies, University of West Bohemia, Plzen, Czech Republic

Abstract

Recent advances in Artificial Intelligence (AI) have paved the way for the integration of text-based and voice-enabled chatbots as adaptive virtual tutors in education. Despite the increasing use of AI-powered chatbots in language learning, there is a lack of studies exploring the attitudes and perceptions of teachers and students towards these intelligent tutors. This study aims to compare several linguistic and technological aspects of four App-Integrated Chatbots (AICs) and to examine the perceptions among English as a Foreign Language (EFL) teacher candidates. In this mixed-methods research based on convenience sampling, 237 college students from Spain (n = 155) and the Czech Republic (n = 82) interacted with four AICs over a month, and evaluated them following a rubric based on the Chatbot-Human Interaction Satisfaction Model. This scale was specifically designed to assess different linguistic and technological features of AICs such as response interval, semantic coherence, sentence length, and user interface. Quantitative and qualitative data were gathered through a pre-post-survey, based on the CHISM model and student assessment reports. Quantitative data were analyzed using SPSS statistics software, while qualitative data were examined using QDA Miner software, focusing on identifying recurring themes through frequency analysis. The findings indicated a moderate level of satisfaction with AICs, suggesting that enhancements in areas such as better adapting to learner needs, integrating interactive multimedia, and improving speech technologies are necessary for a more human-like user interaction.

Keywords: Chatbots, Education, Language learning, Teacher candidates, Evaluation

Introduction

Chatbot technology has evolved rapidly over the last 60 years, partly thanks to modern advances in Natural Language Processing (NLP) and Machine Learning (ML) and the availability of Large Language Models (LLMs). Today chatbots can understand natural language, respond to user input, and provide feedback in the form of text or audio (text-based and voice-enabled). They can offer learners the possibility to engage in simulated conversational interactions in a non-judgmental environment (El Shazly, 2021; Skjuve et al., 2021). For these reasons, chatbots are being increasingly used as virtual tutors to facilitate the development of language skills and communicative

competence in the target language (Huang et al., 2022; Hwang & Chang, 2021; Zhang et al., 2023).

Additionally, chatbots can be employed to furnish language learners with supplementary resources and provide immediate assistance such as access to online dictionaries, digital materials, and social media in real-time (Dokukina & Gumanova, 2020; Haristiani & Rifa'i, 2020). Modern chatbots can include speech technologies (Recognition and Synthesis or R&S) and can be customized to cater to the specific needs of individual learners, thus allowing for the provision of personalized feedback and adaptive support in the learning process (Jeon et al., 2023; Kim et al., 2019).

Thanks to these advances, the incorporation of chatbots into language learning applications has been on the rise in recent years (Fryer et al., 2020; Godwin-Jones, 2022; Kohnke, 2023). Their main goal is to engage learners in simulated conversations in a digital environment, providing interactive exercises to practice pronunciation, vocabulary, and grammar, detecting and correcting errors in real-time, and adapting the instruction to the individual learner's needs. The wide accessibility of chatbots as virtual language tutors, regardless of temporal and spatial constraints, represents a substantial advantage over human instructors.

However, previous studies (Huang et al., 2022; Kim et al., 2019; Vera, 2023) have identified several limitations associated with the adoption of chatbots in language instruction such as the presence of redundancies (off-topics and prearranged answers), limited ability to understand more complex questions or sentence structures (sentence length and complexity), limited understanding of the contextual meaning and non-standard language (lexical richness), and the inability to engage in multiuser conversational interactions. These limitations may result in learners' lack of engagement and satisfaction with the chatbot (Jeon, 2021). Nevertheless, these limitations often stem from the chatbot design and are not inherent flaws. As AI technology is rapidly evolving, some of these issues are being progressively addressed, enhancing chatbots' capabilities and potentially reducing learners' dissatisfaction over time.

Research on app-integrated chatbots (AICs henceforth) in language learning is relatively scarce compared to the extensive literature on mobile app usage in language learning (MALL), primarily due to the emerging nature of this AI technology. Consequently, the present study aims to investigate several linguistic and technological features of four AICs (*Mondly, Andy, John Bot, and Buddy.ai)* and to examine teacher candidates' perceptions toward them through an ad-hoc model named the Chatbot-Human Interaction Satisfaction Model (CHISM). The novelty of this research is twofold. Firstly, it aims to analyse EFL teacher candidates' perceptions and interest in AICs as current students and future educators. Secondly, it proposes an adapted model of CHISM (Belda-Medina et al., 2022) to examine learner satisfaction with AICs, considering both technological and linguistic dimensions, which can be applied to future research. Our study specifically addresses the limitations previously mentioned in chatbot integration for language learning by focusing on the user's experience from three distinct perspectives: language, design, and user interaction. Through the application of the CHISM model, we comprehensively evaluate these aspects, setting the foundation for future research in language education.

The study has three main objectives. Firstly, it aims to investigate the current knowledge and opinions of language teacher candidates regarding App-Integrated Chatbots (AICs). Secondly, it seeks to measure their level of satisfaction with four specific AICs after a 1-month intervention. Lastly, it aims to evaluate their perspectives on the potential advantages and drawbacks of AICs in language learning as future educators.

Literature review

Chatbot definition

A chatbot, short for chatterbot, is a computer program that uses artificial intelligence (AI) to conduct a conversation via auditory or textual methods and interacts with humans in their natural languages. These interactions usually occur through websites, messaging applications, or mobile apps, where the bot is capable of simulating and maintaining human-like conversations and perform different tasks (Adamopoulou & Moussiades, 2020).

In this research, the term chatbot (AIC) is used to refer to virtual tutors integrated into mobile applications specifically designed for language learning to provide students with a personalized and interactive experience. These AICs may cover different aspects of language learning, such as grammar, vocabulary, pronunciation, and listening comprehension, and use various techniques to adapt to the user's level of proficiency and tailor their responses accordingly.

App-Integrated Chatbots (AICs) in language learning

The proliferation of smartphones in the late 2000s led to the integration of educational chatbots into mobile applications. However, the initial models were basic, relying on a scripted question–answer format and not intended for meaningful practice beyond their specific subject area (Godwin-Jones, 2022). Since then, AI technology has significantly advanced and chatbots are now able to provide more comprehensive language learning support, such as conversational exchange, interactive activities, and multimedia content (Jung, 2019; Li et al., 2022).

Existing literature on AIC integration focuses on three main areas. The first one delves into the effects of AICs on language competence and skills. Research in this area probes the efficacy of chatbots in fostering significant improvements in different linguistic aspects, including but not limited to grammar (Haristiani et al., 2019; Kharis et al., 2022; Kim, 2019), vocabulary (Ajisoko, 2020; Kim, 2020), writing (Pin-Chuan Lin & Chang, 2020), and conversation (Hakim & Rima, 2022; Pham et al., 2018). These studies showed how AICs can manage personal queries, correct language mistakes, and offer linguistic support in real-time. However, some authors noted certain limitations, such as their reliance on large data for learning and prediction, and their potential inability to understand different accents, language nuances, and context, which could lead to conversational errors (Huang et al., 2022; Panesar, 2020).

.The second major theme is the impact of AICs on learner engagement and motivation (Dokukina & Gumanova, 2020; Li et al., 2022). This line of research investigates how the interactive nature of some AICs can reduce students' anxiety and cognitive load (Hsu et al., 2021) and promote an engaging learning environment (Bao, 2019). Furthermore, some authors have examined the ability of chatbots to promote self-directed

learning, given their wide availability and capacity for personalized responses (Annamalai et al., 2023). Nonetheless, certain researchers, including Ayedoun et al. (2015) and Fryer et al. (2019), have indicated that the initial enthusiasm and engagement students show towards chatbots may be short-lived, attributing this to the novelty effect of this technology.

The third area explores how AICs' design can positively affect language learning outcomes. Modern AICs usually include an interface with multimedia content, real-time feedback, and social media integration (Haristiani & Rifa'I, 2020). They also employ advanced speech technologies to ensure accessible and humanlike dialogues (Petrović & Jovanović, 2021). Additionally, AICs today can also incorporate emerging technologies like AR and VR, and gamification elements, to enhance learner motivation and engagement (Kim et al., 2019). However, some studies have also noted certain limitations, such as repetitive exercises and imperfect speech technologies, suggesting they should be used only as supplementary tools (Jung, 2019) or assistants (Kukulska-Hulme & Lee, 2020) rather than intelligent virtual tutors.

Teachers and learners' views on the use of AICs for language learning

The landscape of mobile-application language learning (MALL) has been significantly reshaped in recent years with the incorporation of AICs (Pham et al., 2018). This innovative approach to mobile learning has been positively received by both students and teachers. For example, Chen et al. (2020) highlighted the effectiveness of AICs for Chinese vocabulary learning by comparing chatbot-based tutoring with traditional classroom settings. The study reported positive user feedback on the chatbot's ease of use, usefulness, and enjoyment, as measured by the Technology Acceptance Model (TAM). Similarly, Yang (2022) underscored the favourable views of AICs in English language education, with teachers valuing the chatbot's capacity to manage routine tasks, thereby allowing them to concentrate on more substantial classroom duties. In this study, students appreciated the supplemental use of chatbots for their ability to provide immediate feedback on unfamiliar words or concepts, thereby enriching their English textbook learning.

However, the use of AICs as virtual tutors also presents certain challenges. Some studies have emphasized that interactions with AICs can seem detached and lack the human element (Rapp et al., 2021). Additionally, while AICs can handle a wide range of queries, they may struggle with complex language nuances, which could potentially lead to misunderstandings or incorrect language usage. It has also been observed that some students' interest dwindled after the initial period of engagement due to repetitive conversation patterns and redundancies, making the interaction less natural compared to student–teacher exchanges (Fryer et al., 2019).

In our study, the term 'perceptions' is defined, following Chuah and Kabilan's approach (2021), as users' attitudes and opinions towards their interactions with chatbots in education. This encompasses aspects such as perceived usefulness, acceptance, and potential interest. Research in this area underscores the importance of understanding users' viewpoints on chatbots, including their acceptance of these tools in educational settings and their preferences for chatbot-human communication. Similarly, 'satisfaction' is described as the degree to which users feel that their needs and expectations are met

by the chatbot experience, encompassing both linguistic and design aspects. Studies like those by Chocarro et al. (2023) have delved into students' enjoyment and engagement with chatbots, highlighting the importance of bot proactiveness and individual user characteristics in shaping students' satisfaction with chatbots in educational settings.

Despite these insights, there remains a significant gap in the literature regarding a comprehensive understanding of teachers' and students' perceptions of AICs, particularly in how these perceptions influence their acceptance and effectiveness in language education. This gap is more pronounced in understanding how the design and linguistic features of AICs impact user satisfaction and engagement. While studies like those of Chen et al. (2020) and Chocarro et al. (2023) have begun exploring these areas, there is a need for a more targeted framework to evaluate satisfaction with AICs in the context of language learning. To address this need, our study investigates EFL teacher candidates' levels of satisfaction and perceptions of four AICs. We propose the Chatbot-Human Interaction Satisfaction Model (CHISM), an adaptation from a previous model used with intelligent conversational agents (Belda-Medina et al., 2022), to specifically measure and analyze these perceptions and their impact on language learning among language teacher candidates.

Questions and objectives

The study is structured around three research questions:

- What prior knowledge do teacher candidates have about using App-Integrated Chatbots (AICs) in language learning?
- How satisfied are participants with the four AICs selected after a 1-month intervention?
- What potential benefits and limitations of AICs do teacher candidates perceive in language learning?

These questions align with three objectives. The first objective explores the existing knowledge and understanding of language teacher candidates about AICs in language learning. This includes their familiarity with chatbot technologies, perceptions of their utility, and any prior experiences they might have with these tools. The second objective assesses participants' satisfaction levels with four selected AICs after a 1-month intervention, utilizing the Chatbot-Human Language Interaction Satisfaction Model (CHISM) to evaluate linguistic accuracy and user interface design. The third objective evaluates the participants' perceptions as future educators of the potential benefits and limitations of AICs in language learning. This objective considers their views on how AICs might be integrated into future language education settings, highlighting their potential impact on teaching and learning methodologies. The corresponding research objectives are as follows:

- O1. Explore the prior knowledge of language teacher candidates regarding chatbots.
- O2. Measure their level of satisfaction after the intervention with four AICs through the Chatbot-Human Language Interaction Satisfaction Model (CHISM).

- O3. Evaluate the participants' perceptions of AICs in language learning as future educators.

Method

Participants and context

The research, conducted over two academic years (2020–2022) with a mixed-methods approach and convenience sampling, initially involved 163 students from the University of X (Spain) and 86 from the University of X (Czech Republic). However, the final participant count was 155 Spanish students and 82 Czech students, as some declined to participate or did not submit the required tasks. Participation was voluntary, and students who actively engaged with the chatbots and completed all tasks, including submitting transcripts and multiple-date screenshots, were rewarded with extra credits in their monthly quizzes. This approach ensured higher participation and meaningful interaction with the chatbots, contributing to the study's insights into the effectiveness of AICs in language education.

Participants were third-year-college students enrolled in two subjects on Applied Linguistics taught over the course of 4 months, with two-hour sessions being held twice a week. Both Applied Linguistics courses are integral components of the Teacher Education degree programs at the respective universities in Spain and the Czech Republic. These participants were being trained to become English language teachers, and the learning module on chatbot integration into language learning was strategically incorporated into the syllabus of both subjects, taught by the researchers. The choice of Spain and the Czech Republic was primarily based on convenience sampling. The two researchers involved in this study are also lecturers at universities in these respective countries, which facilitated access to a suitable participant pool. Additionally, the decision to include these two different educational settings aimed to test the applicability and effectiveness of AICs across varied contexts. The study found similar results in both settings, strengthening the argument for the broader relevance and potential of AICs in diverse educational environments.

The language proficiency of the students aligned with the upper intermediate (B2) and advanced (C1) levels as defined by the Common European Framework of Reference for Languages (CEFR), while some participants were at the native speaker (C2) level. In our study, the primary focus was on evaluating language teacher candidates' perceptions of AICs in language learning, rather than assessing language learning outcomes. Considering that the majority of participants possessed an upper intermediate (B2-C1) or advanced (C2) proficiency level, the distinction between native and non-native speakers was not deemed a crucial factor for this research. Subsequently, a statistical analysis was conducted to evaluate the impact of language nativeness (Spanish and Czech versus non-Spanish and non-Czech speakers), revealing no significant differences in the study's outcomes. Furthermore, the evaluations of the AICs by both Spanish and Czech cohorts displayed similar results. This analysis led us to conclude that language nativeness and the specific educational settings of the participants were not key factors influencing the results of our study. Regarding gender, 81% of the participants were females, while 19% were male students. All participants were under 30 years of age.

The research was carried out following the regulations set by each institution for interventions with human subjects, as approved by their respective Ethical Committees. Participants provided written consent for the publication of their interactions with chatbots for academic purposes. All data obtained were anonymised and analysed confidentially.

Instruments and procedure

Following a sequential QUAN-QUAL approach (Nastasi et al., 2007), data were gathered through a pre-post survey, class discussion and assessment reports. The pre-survey (15 items) was divided into two sections: a socio-demographic Section (3 items), and a technological affinity Section (12 items) that focused on participants' usage, prior knowledge, and perceptions toward chatbots. The post-survey (15 items) was designed to gather data related to the use of the four AICs (*Mondly, Andy, John Bot, Buddy.ai*) employed in the intervention. The study incorporated th.

The Chatbot-Human Interaction Satisfaction Model (CHISM) is a tool previously designed and used to measure participants' satisfaction with intelligent conversational agents in language learning (Belda-Medina et al., 2022). This model was specifically adapted for this study to be implemented with AICs. The pre-post surveys were completed in the classroom in an electronic format during class time to ensure a focused environment for the participants. Quantitative data obtained were analysed using the IBM® SPSS® Statistics software 27. The main objective was to determine the average responses by calculating the means, evaluate the variability in the data by measuring the standard deviation, and assess the distribution's flatness through kurtosis.

Qualitative data were collected through class discussions and assessment reports of the AICS following a template provided through the Moodle platform. During the 1-month intervention period in each educational setting, participants independently completed the assessment reports. They were instructed to provide personal feedback on their interaction with each AIC, using the template to note both positive and negative aspects. Additionally, they were asked to attach 12 screenshots illustrating their interaction, three with each AIC, to support their assessment. QDA Miner Software was used for textual analysis of students' written evaluations on each AIC, adhering to a provided template. Student comments were systematically categorized into potential benefits and limitations following the template structure and then coded using a tree-structured code system, focusing on recurrent themes through frequency analysis. The research comprised four stages as shown in Fig. 1.

For the interaction, detailed instructions were provided via Moodle, with the aim not to measure the participants' English learning progress, but to enable critical analysis of each AIC as future educators. The teacher candidates were guided on how to engage with the chatbots, including selecting different language levels, using varied sentence types, introducing typical errors, exploring voice options, and investigating the use of AR and other technologies if available. This assessment was aligned with the CHISM scale, which was completed in a post-survey. A minimum interaction of three hours per week with each AIC, or 48 h over a month across all AICs, was requested from each participant.

The selection of the four AICs, namely *Mondly, Andy, John Bot, and Buddy.ai,* was guided by specific criteria, including multiplatform compatibility, wide availability, and

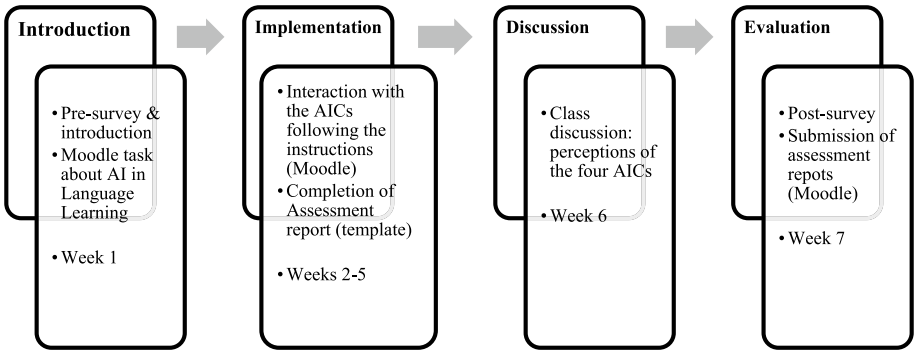

Fig. 1 Research procedure

Table 1 Description of the four AICs used in the study (updated on 11/07/2023)

AICs	Mondly	Andy	John Bot	Buddy.ai
Company or founder	ATi Studios (Romania)	ZTO labs (Russia)	Buabit (Sweden)	AI Buddy Inc. (USA)
Release	2014	2016	2020	2017
Platform	iOS/Android	iOS/Android	iOS/Android	iOS/Android
Language/s available	41, English included	English	English	English
Pricing	6 lessons free $9.99/ month	Free Premium	Free Premium	Free trial $4.99/month
website	www.mondly.com/	andychatbot.com	buabit.com	buddy.ai/en
Special features	Voice-enabled, AR & VR	Voice-enabled, interactive testing	Voice-enabled, translation options	voice-based, Mixed Reality, Children-oriented (4–10),

diverse functionalities such as the integration of different technologies. These AICs offered a wide range of options, such as catering to different English language proficiency levels, providing personalized feedback, adapting to individual learning progress, and incorporating other technologies (AR, VR) in some cases. The aim was not to compare the four AICs, but rather to present teacher candidates with a broad overview of these virtual tutors, providing a variety of options and examples. Table 1 summarizes the main features of each AIC.

Results

Concerning general perceptions of chatbots, the pre-survey results shown in Table 2 evidenced that participants found them most useful in providing general information (M = 3.4), while the perceived usefulness was lower for social media (M = 2.7), and education (M = 2.7). Analysis of the data indicated a uniform perspective among Spanish and Czech students on chatbots' role across general information, social media, and education. Mean scores in these areas were closely matched, highlighting a consensus on the significance of chatbots in these domains.

Regarding the frequency of use of the four AICs employed in the intervention, the post-survey results shown in Table 3 indicated that Andy was the most frequently used, averaging nearly 4 h per week, followed by John Bot and Mondly, while Buddy.ai was the least used. Concerning the educational setting, Spanish participants interacted more

Table 2 Perceptions of chatbot usefulness (1 = not useful at all to 5 = extremely useful)

Chatbot usefulness	Total (n = 238)	SD	Kurtosis	Spanish (n = 155)	Czech (n = 82)
	M			M	M
General information	3.44	1.765	−,384	3.41	3.49
Social media	2.70	1.711	−1,172	2.76	2.57
Education	2.71	1.698	−1,119	2.68	2.76

Table 3 Frequency of interaction

Frequency of use	Total (n = 238)	SD	Spanish (n = 155)	Czech (n = 82)
	M		M	M
Mondly	1.71	0.679	1.78	1.56
Andy	1.83	0.693	1.89	1.73
John Bot	1.75	0.676	1.76	1.71
Buddy.ai	1.34	0.606	1.44	1.13

Values: 1 (3 h approx./week), 2 (4 h approx. week), 3 (5 + hours/week)

Table 4 CHISM results: Language Experience (LEX) dimension

α = .832 Item	Abbr	Mondly M	SD	Andy M	SD	John M	Bot SD	Buddy M	.ai SD
#1	SCB	2.54	1.169	2.89	1.010	2.78	1.138	2.33	1.326
#2	SLC	2.68	1.102	3.07	1.055	2.92	1.073	2.15	1.269
#3	R&S	2.87	1.277	2.29	1.237	2.71	1.217	3.10	1.489
#4	QVR	2.92	1.223	3.18	1.042	3.08	1.216	2.60	1.389
#5	QGE	2.47	1.124	2.90	1.239	2.81	1.169	2.10	1.297
#6	EDC	2.16	1.098	2.46	1.149	2.49	1.066	2.15	1.216
#7	NCI	2.37	1.124	2.84	1.135	2.64	1.167	2.24	1.275
#8	CRI	3.11	1.279	3.45	1.196	2.99	1.212	2.76	1.494
#9	NVL	2.47	1.255	2.90	1.325	2.78	1.220	2.19	1.576

Based on a 5-point Likert scale (1 = not satisfied at all to 5 = completely satisfied)

frequently with all four AICs compared to Czech students. The SD values show a similar level of variation in the weekly interaction hours across all four AICs for both Spanish and Czech participants, suggesting a comparable spread of interaction frequencies within each group.

Chatbot-human interaction satisfaction model results

The CHISM scale comprised three dimensions: Language Experience (LEX), Design Experience (DEX) and User Experience (UEX). Table 4 shows the results of the first dimension (LEX) aimed at measuring nine language-related features of the four AICs.

Semantic Coherent Behaviour (#1SCB) refers to the chatbot's capacity to sustain a contextually relevant and meaningful dialogue with the user. Prior research has evidenced that student dissatisfaction with chatbot interaction is mainly due to the use of prearranged responses and off-topic remarks (Fryer et al., 2020; Kukulska-Hulme & Lee,

2020). Therefore, it is crucial for a chatbot to comprehend the conversation's context, provide appropriate responses, and recall past interactions. None of the AICs reached the moderate point of 3 on a five-point Likert scale, with Andy scoring the highest (M = 2.8), and closely followed by John Bot (M = 2.7). These scores can be explained by considering each AIC's design and target audience. For instance, Mondly heavily depends on pre-programmed responses as 'it is targeted at lower levels' (Hajizadeh et al., 2023: 12), requiring students to select a given response from a limited set of options. As a result, participants deemed the interaction as repetitive, because 'if learners do not follow the assumed conversation patterns, the chatbot repeats the same questions until the learner provides the expected answer' (Jung, 2019: 77). Similarly, Buddy.ai is primarily designed for children aged 4–10 with a focus on oral skills through repetition drills. While these iterative approaches can ensure response accuracy and consistency at lower levels, they may restrict the chatbot's capability to engage in more dynamic and contextually relevant conversations, as pointed out in previous research (Gokturk, 2017).

Sentence Length and Complexity (#2SLC) pertains to the structure and variety of the sentences that a chatbot uses in its responses. Ideally, an AIC playing the role of a virtual tutor should adjust the length of its responses based on the learner's level, inputs and the context of the conversation, including the use of different grammatical constructs and vocabulary. Conversely, an AIC that only uses simple, repetitive patterns might come off as robotic or limited. Among the four AICs, Andy was perceived as the most proficient in varying sentence length and complexity depending on the learner's inputs (M = 3.0). Unlike Mondly and Buddy.ai where the chatbots' structure is limited according to the students' lower levels, 'Andy bot commands a high level of vocabulary and sentence structures and can deal with difficult topics; thus it can accommodate advanced learners with more free conversation style' (Jung, 2019: 79). As Andy is a conversation-oriented rather than a lesson-oriented chatbot, it can cater to different proficiency levels, from beginners to advanced, so it dynamically adapts to the individual language level of each learner.

The third item responds to Speech Recognition and Synthesis (#3R&S), which has received limited attention in relevant research (Jeon et al., 2023). Early chatbots relied on simple text-based inputs and outputs, and their ability to understand and respond to user questions was quite limited. Recognition technology, often referred to as Automatic Speech Recognition (ASR), converts spoken language into written text while Synthesis technology, also known as Text-to-Speech (TTS), converts written text into spoken words. R&S technology allows AICs to provide text and verbal responses to user inputs, making the interaction more engaging and human-like. Buddy.ai, with its emphasis on oral interaction for children, outperformed the other chatbots in this feature as the results in Table 4 show, while some users reported speech recognition problems with the other AICs, necessitating multiple repetitive attempts at interaction. These technical issues, partly due to an intentional accented pronunciation, made some participants feel anxious about their utterances and lowered their motivation to interact with the chatbots, in line with previous findings (Jeon, 2021). Additionally, the audio quality of certain AICs was criticized for their 'robotic' sound, as illustrated in the qualitative results presented in "Teacher candidates' perceptions of App-Integrated Chatbots" section (Table 7). While R&S technology has come a long way, there are still several challenges

that need to be addressed to improve the quality of chatbot-human interactions: accurate speech recognition, emotion recognition (intonation, pitch, rhythm), and natural language generation.

Items 4 and 5 are associated with the Quality of Vocabulary Reference (#4QVR) and Quality of Grammar Explanations (#5QGE) respectively. These items, which have been incorporated into the original CHISM scale to be utilized specifically with AICs, are essential for learners to comprehend word usage in context (synonyms, collocations) and understand grammar rules. Among the four AICs, Andy received the highest scores in both #4QVR (M = 3.1) and #5QGE (M = 2.9), indicating its higher effectiveness in providing comprehensive vocabulary and grammar references, closely followed by John Bot (#4QVR M = 3.0 and #5QGE M = 2.8). Both chatbots prioritize context-based explanations, offering learners the flexibility to personalize their learning progress (adaptive learning) in real time and engage with vocabulary and grammar rules in different contexts, as Fig. 2 illustrates. Furthermore, the integration of vocabulary and grammar exercises as reinforcement to the provided explanations played a crucial role in more extensive human-chatbot interaction, in line with previous findings (Jung, 2019).

Error detection and correction (#6EDC) is another essential component of AICs as it plays a significant role in language learning. It enables the chatbot to identify and provide feedback on language errors in real time, helping learners improve their accuracy while fostering a supportive learning environment. Prior research has indicated that the

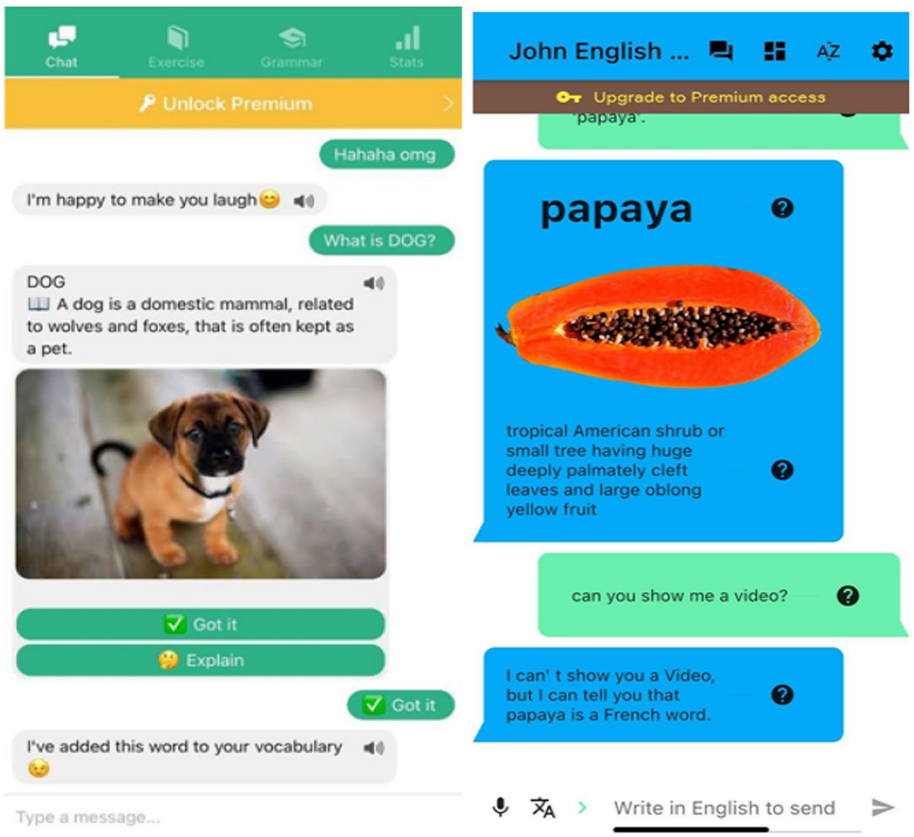

Fig. 2 Screenshots of the participants' interaction with Andy (left) and John Bot (right)

use of AICs can reduce language anxiety, especially among less self-confident learners, by creating a non-judgmental environment where learners feel less inhibited to make errors and participate in conversations (Bao, 2019; Huang et al., 2022). However, despite previous findings highlighting the positive use of John Bot in this aspect (Khang et al., 2023), the EDC results shown in Table 4 indicated that none of the AICs achieved a moderate level of satisfaction. The scores were slightly higher for Andy (M = 2.4) and John Bot (M = −2.4), though. This can be attributed to two factors: first, the limited range of response options given in the interaction with some AICs, which restricted their ability to effectively identify and correct errors in a more natural conversation, as already pointed out by Jung (2019); and second, the lack of meaningful explanations provided by certain AICs, particularly those that are more lesson-oriented. Participants noted that Mondly and Buddy.ai lacked clear explanations to help learners understand and correct their errors. As pointed out by teacher candidates, the practice of merely asking for repetition of the same answers without further explanation could hinder the learning process, suggesting a need for improvements in this area.

Natural Conversational Interaction (#7NCI) pertains to the chatbot's ability to emulate the natural flow and dynamics of human conversation. It involves several key elements, such as maintaining a contextually relevant conversation, understanding and responding appropriately to user inputs, demonstrating empathy, and adapting the language style and tone to suit the learner's preferences. The goal is to create a conversation that not only provides informative and accurate responses but also engages users in a manner that simulates a human-to-human interaction. None of the AICs reached the desired level of conversational naturalness, as participants found their responses predictable and lacking the adaptability seen in human tutors. As observed in previous research (Kukulska-Hulme & Lee, 2020), the limited range of response options and the inability of AICs to provide personalized feedback and multi-user interaction like a human tutor contributed to the perception that AICs fall short in simulating human tutoring experience.

Chatbot Response Interval (#8CRI) relates to the time it takes for a chatbot to generate a response following text or voice-based input from the learner. During chatbot interactions, there can be a delay in response due to factors such as processing time and system limitations as AICs often need additional time to analyse input, retrieve information, and generate suitable responses. The chatbot response time, whether overly rapid or delayed, can shape the overall dialogue experience and affect how engaged the learner is and how human-like they perceive the chatbot to be (Gnewuch et al., 2022). Several teacher candidates reported technical issues, particularly with the voice-enabled features of some AICs, as certain responses took longer or were not accurately processed. Among the four AICs, Andy demonstrated a higher score in #8CRI (M = 3.4), indicating a prompt and adaptive response to user inputs while Buddy.ai faced challenges in this aspect, particularly related to speech recognition problems, resulting in a lower score (M = 2.7).

Non-Verbal Language (#9NVL) in written communication refers to the use of visual cues, gestures, and symbols to convey meaning and emotions in the absence of spoken words. Emojis, memes, stickers, GIFs, and other visual representations are fundamental in social media and messaging apps today, yet their application in chatbots is an area that remains under-researched. These elements enable individuals to

communicate more effectively, evoke specific emotions, and enhance the overall user experience (Beattie et al., 2020). Furthermore, NVL compensates for the limitations of text-based communication by filling the gap left by the absence of tone of voice, facial expressions, and body language. In the context of AICs, the integration of NVL can foster more engaging and expressive conversations, nurture a sense of community, and promote better understanding among learners. Based on the data, the teacher candidates perceived all AICs to have limited effectiveness in using visual cues and symbols to enhance communication and convey emotions, resulting in low results in NVL except for Andy (M = 2.9), which scored higher due to its more frequent use of these elements as illustrated in Fig. 2.

The second dimension of the CHISM model, focusing on the Design Experience (DEX), underscores its critical role in fostering user engagement and satisfaction beyond the linguistic dimension. Elements such as the chatbot interface and multimedia content hold substantial importance in this regard. An intuitive and user-friendly interface enriches the overall user experience and encourages interaction (Chocarro et al., 2021; Yang, 2022). Additionally, the incorporation of engaging multimedia content, including videos, images, and other emerging technologies, can also increase users' attention and engagement (Jang et al., 2021; Kim et al., 2019). Table 5 shows the results of the three items included in the DEX dimension.

The Multimedia Content design (MC #10) encompasses the use of images and videos to create a multimedia-rich environment that enhances language learning (Zhang et al., 2023). Furthermore, AICs can be integrated with social media platforms, promoting collaboration and cultural exchange (Haristiani & Rifa'i, 2020). The integration of emerging technologies such as VR and AR further enhances the language learning experience, providing immersive and authentic learning environments. The results of the four AICs indicated that Buddy.ai achieved the highest score (M = 3.5), closely followed by Mondly (M = 3.1). The teacher candidates enjoyed Buddy.ai's innovative design as it specifically caters to children and incorporates mixed reality elements, so it creates a more engaging and immersive learning environment for young learners as Fig. 2 illustrates. Similarly, Mondly's AR component was praised as innovative, offering interactive lessons and realistic conversations for vocabulary learning and pronunciation practice although some participants considered AR more of a novelty than an effective learning tool.

The Game-Based Learning (#11GBL) component refers to the incorporation of challenges, rewards, progress bars, status indicators, and other game-based elements, to enhance learner motivation, track progress, and provide valuable feedback (Jung, 2019; Petrović & Jovanović, 2021). Based on the data, all four AICs received relatively

Table 5 CHISM results: Design Experience (DEX) dimension

α = .861 Item	Abbr	Mondly M	SD	Andy M	SD	John Bot M	SD	Buddy.ai M	SD
10	MC	3.11	1.275	2.37	1.225	2.61	1.217	3.54	1.656
11	GBL	2.95	1.133	3.09	1.075	2.81	1.141	3.68	1.414
12	CUI	3.58	1.187	3.55	1.108	3.34	1.186	3.61	1.642

Based on a 5-point Likert scale (1: not satisfied at all to 5: completely satisfied)

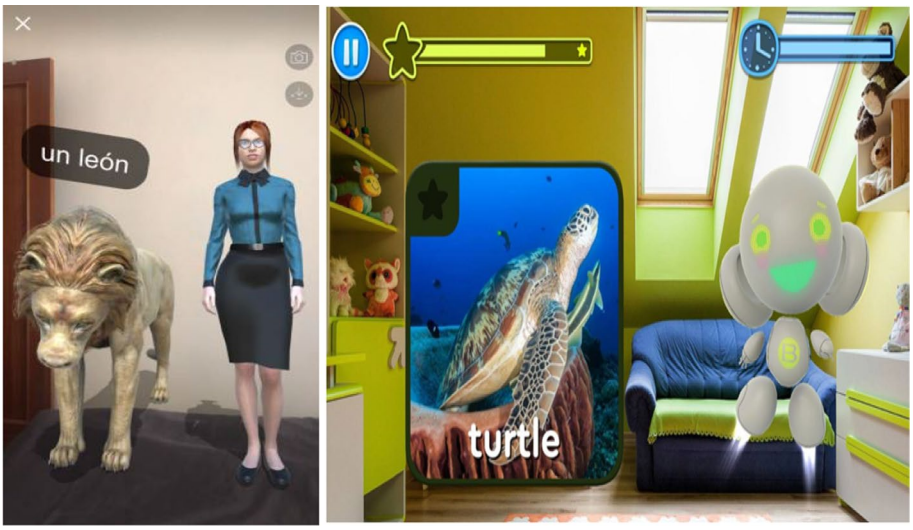

Fig. 3 Screenshots of Mondly AR (left) and Buddy.ai (right)

Table 6 CHISM results: User Experience (UEX) dimension

α = .826 Item	Abbr	Mondly M	SD	Andy M	SD	John Bot M	SD	Buddy.ai M	SD
13	UENG	2.44	1.119	2.89	1.178	2.57	1.174	2.76	1.457
14	UENJ	2.42	1.177	2.99	1.151	2.51	1.269	2.91	1.510
15	UFI	2.41	1.228	2.88	1.282	2.38	1.270	2.84	1.376

Based on a 5-point Likert scale (1: not satisfied at all to 5: completely satisfied)

positive results in terms of GBL with the highest score for Buddy.ai (M = 3.6), indicating its strong performance in incorporating game-based elements for children while talking with an interactive virtual tutor as Fig. 3 illustrates.

A Chatbot User Interface (#9CUI) that is intuitive, scalable, and provides easy access to a variety of multimedia materials may determine its success among language learners (Chocarro et al., 2023; Kim et al., 2019). Some chatbots even offer customization based on the user's profile, learning objectives, and preferences (Belda-Medina et al., 2022). In this sense, all AICs garnered positive feedback from teacher candidates, with Buddy.ai again achieving the highest average rating (M = 3.6), making it well-suited for children. As expressed by several teacher candidates, the importance of a dynamic and customizable CUI is on par with the linguistic abilities of the chatbot, as it affects their engagement and motivation. The results for each of the three items comprising the third CHISM dimension, User Experience (UEX), are presented in Table 6.

The User Engagement (#13UENG) process involves several stages: initiation of interaction by learners, active involvement in tasks or actions, pausing or stopping which indicates disengagement, and re-engagement when learners return to the activity. In order to sustain learner engagement, AICs need to offer memory capacity, immediate assistance, customized experiences, and scalable options. Despite prior research highlighting the positive results in the use of Mondly (Hajizadeh et al., 2023) and John Bot (Khang et al., 2023), teacher candidates reported varying degrees of engagement with

the four AICs, with the overall feedback not being as positive as expected. Both Mondly (M = 2.4) and John Bot (M = 2.5) were perceived as repetitive, while Andy (M = 2.8) was rated more favourably due to its memory capacity, adaptability, and extensive vocabulary range. Participants believed these characteristics brought Andy closer to a virtual intelligent tutor.

User Enjoyment (#14UENJ) pertains to the level of pleasure and satisfaction experienced by learners while interacting with the AIC (Chocarro et al., 2021; Yang, 2022). Learners' satisfaction can be shaped by engaging gamified activities, interactive conversational dialogues, visually appealing designs, motivating progress indicators, and immersive audio effects that collectively contribute to an enjoyable and enriching learning environment. However, the reported satisfaction levels were not as positive as expected in line with the #14UENG results. Mondly received the lowest score (M = 2.4), while Andy (M = 2.9) and Buddy.ai (M = 2.9) attained a moderate level of enjoyment, indicating that their incorporation of game elements and visually captivating aspects contributed to a more satisfying experience among teacher candidates.

User Further Interest (UFI #15) can be defined as the level of curiosity and eagerness displayed by students to explore and learn more about the AIC. It reflects users' willingness to delve deeper into the AIC's capabilities, engage with new functionalities, and continue using it in the future. UFI may be influenced by factors such as perceived value, continuous learning opportunities, and the incorporation of user feedback and adaptability. Users are more likely to be further interested in an AIC that they perceive as valuable and provides on-going learning opportunities (Fryer et al., 2019). The results obtained for Andy (M = 2.8) and Buddy.ai (M = 2.8) were moderate, while Mondly (M = 2.4) and John Bot (M = 2.3) received lower scores, aligning with the findings of the other two components of the UEX dimension.

Teacher candidates' perceptions of app-integrated chatbots

Qualitative data, obtained from in-class discussions and assessment reports submitted through the Moodle platform, were systematically coded and categorized using QDA Miner. The goal was to analyse and identify the main benefits and drawbacks of each AIC as perceived by teacher candidates. These themes were cross-referenced with the different components of the CHISM model to establish correlations as shown in Table 7. Frequency in the table refers to the number of observations made in the sample of textual data based on the written assessments provided by participants.

In line with previous research, the main advantages of AICs as perceived by participants were providing language support and delivering feedback on the learning progress via conversational interaction, in line with previous findings (Jeon, 2021; Yang, 2022), as well as enhancing user involvement through multimedia and interactive activities (Chuah & Kabilan, 2021; Dokukina & Gumanova, 2020). Conversely, the key limitations highlighted were their lack of adaptivity to varying proficiency levels (Huang et al., 2022), a tendency of some AICs to provide unrelated responses due to an over-reliance on predetermined answers (Huang et al., 2022), and notably, the imperfections in voice recognition capabilities (Bao, 2019). Table 7 provides a summary of the primary advantages and drawbacks of each AIC, along with their correlation to the items in the CHISM model, which are indicated in parentheses.

Table 7 Qualitative analysis results using QDA Miner

AIC	Categ	Codes (CHISM Item/s)	Freq. (%)
Mondly	Benefits	Provides support with basic conversations, suggests different answers to questions, provides options in case you don't know what to answer (QVR, QGE)	64.5
		options to practice dialogues, vocabulary, and take tests, recording functions, option to speak to the AIC (QID, R&S)	53.0
	Limitations	Limiting answers to only a few words, limited conversation topics, it can be repetitive, tells you what to answer, not challenging, (SLC, NCI, ENJ, ENG)	73.5
		Voice recognition feature may not always work properly, it can be frustrating for users who want to practice their speaking skills, not good at detecting different accents, very sensitive to background noise (R&S)	42.5
Andy	Benefits	Rich vocabulary, responses were very accurate and helpful, it provided useful feedback on language skills, good ability to respond in multiple languages (SCB, SLC, EDC)	73.0
		Use of emojis made the conversation more engaging and enjoyable; good response interval, it made the conversation feel more natural and authentic (CRI, MC, UENG, UENJ)	61.5
	Limitations	Responses were sometimes too long and complex, which could be overwhelming for beginners (SCL)	46.5
		Basic interface that can make it feel like users are talking to a machine rather than a person (CUI)	38.0
John Bot	Benefits	Provides a lot of topics to choose from, gives default answers in case you don't know what to answer; provides clear and concise explanations of grammar rules and vocabulary; provides personalized feedback and suggestions for improvement (SCB, EDC, QUI, QGE)	71.5
		User-friendly interface, fast response time, use of multimedia, pronunciation feature, ability to respond in multiple languages (CUI, CRI, MC)	59.0
	Limitations	It sometimes provides answers that have no bearing on the user's questions or ignores questions (SCB)	42.0
		Pronunciation feature can be inaccurate (R&S)	40.5
Buddy.ai	Benefits	Use of interactive exercises and quizzes (progress bars), and games (levels and rewards) makes the conversation engaging and interactive (GBL,UENG)	61.0
		Easy-to-use interface; use of real-life scenarios, videos and multimedia makes the conversation more enjoyable (CUI, MC, UENJ)	57.5
	Limitations	Responses can be too slow and sometimes repetitive, which can make the conversation feel boring and less engaging for children (SCB, SCL)	44.5
		Speech recognition feature can be inaccurate, requires several attempts, it can be confusing for children with accented pronunciation (R&S)	41.0

Discussion

The CHISM results, particularly in the Language Experience (LEX) dimension, revealed significant insights about the teacher candidates' perceptions of the four evaluated chatbots. When examining why none of the AICs achieved moderate satisfaction in the LEX dimension, it is crucial to consider each AIC's design and target audience limitations, as pointed out in previous research (Gokturk, 2017; Hajizadeh, 2023). For instance, Mondly's reliance on pre-programmed responses and Buddy.ai's focus on repetitive drills for children limit dynamic conversation, resulting in lower satisfaction in maintaining contextually relevant dialogues. Although Andy scores slightly higher, it still reveals a need for more adaptable conversation styles for advanced learners. The satisfaction levels in the LEX dimension may also depend on the chatbots' design relative to students' levels, with significant differences observed among the four AICs. For example, while Buddy.ai is oriented towards developing oral skills in children at a lower level, John Bot and Andy

are designed for vocabulary and grammar building through role-playing interactions at more intermediate levels.

Additionally, speech technologies emerged as an area requiring substantial improvement, in line with previous results (Jeon et al., 2023). With the exception of Buddy.ai, the voice-based interactions provided very low results due to poor speech recognition and dissatisfaction with the synthesized voice, potentially leading to student anxiety and disengagement. Improvement could be achieved by investing in advanced technology capable of understanding a wide range of accents and mitigating background noise interference, possibly employing machine learning algorithms trained on various accents and speech patterns (Kohnke, 2023; Kukulska-Hulme & Lee, 2020). Simultaneously, rendering the AICs' voice generation more human-like can be attained through more sophisticated Text-to-Speech (TTS) systems that mimic the intonation, rhythm, and stress of natural speech (Jeon et al., 2023).

The findings indicate other key potential areas for AIC improvement to better cater to users' proficiency levels. It would be beneficial to implement more sophisticated AI algorithms capable of effectively assessing a user's language skills based on their real-time input and adjusting the chatbot's responses accordingly, as learners' interest in chatbots is usually mediated by competence level. The development of LLM-power chatbots could help avoid irrelevant responses often resulting from an over-reliance on pre-set answers, as indicated by Jeon (2021).

Expanding on the necessity for improved customization in AICs, the integration of different features can be proposed to enhance chatbot-human personalization (Belda-Medina et al., 2022). These features include the ability to customize avatars (age, gender, voice, etc.) similar to intelligent conversational agents such as Replika. For example, incorporating familiar characters from cartoons or video games into chatbots can enhance engagement, particularly for children who are learning English by interacting with their favorite characters. Furthermore, by incorporating Augmented Reality (AR) technology, avatars can be launched and video calls can be enabled on social platforms such as Kuki.ai, thereby adding a layer of personal interaction. Looking ahead, allowing students to select specific design aspects of AICs, similar to choosing linguistic features such as target level or accent, could be a crucial step in creating a more adaptive and personalized learning experience.

Conclusion and implications

The application of the CHISM model in the evaluation of four AICs has provided valuable insights into the effectiveness of these tools in language learning. The model, which comprises three dimensions (LEX, DEX, UEX), has allowed for a comprehensive assessment of the AICs across multiple facets. The Language Experience dimension (LEX), which includes elements such as Semantic Coherent Behaviour, Sentence Length and Complexity, and Speech Recognition and Synthesis, revealed that none of the AICs reached the moderate point of satisfaction among EFL teacher candidates. This suggests that while these tools have made strides in providing language-related features, there is still room for improvement, particularly in terms of maintaining contextually relevant dialogues and varying sentence complexity based on the learner's level.

The Design Experience dimension (DEX) underscored the importance of user-friendly interfaces and engaging multimedia content in fostering user engagement and satisfaction. The findings uncovered the necessity for enhancements in adaptive user interfaces, as well as the incorporation of social media and emerging technologies, to simulate the human-student interaction and enrich the language learning experience. The User Experience dimension (UEX) revealed that while some AICs were able to provide a moderate level of enjoyment and engagement, overall satisfaction levels were not as positive as expected. This indicates the need for AICs to offer a more personalized learning experience to sustain learner engagement and interest.

The CHISM model offers a comprehensive approach to evaluating AICs, encompassing not only linguistic capabilities but also design and user experience aspects. This holistic evaluation allows for a more nuanced understanding of the strengths and weaknesses of AICs, providing valuable insights for future improvements. The model also highlights the potential of AICs in language learning, particularly in terms of providing immediate feedback, and fostering a supportive learning environment.

However, the study also highlights the challenges that need to be addressed, such as the requirement for more sophisticated AI algorithms capable of adjusting to the learner's proficiency level and the improvement of speech technologies. This suggests the need for evolving teaching methods and curricula to more effectively incorporate AICs, emphasizing the enhancement of their capabilities for providing contextually rich and varied linguistic experiences. One practical approach could be the introduction of specific learning modules on different types of chatbots, such as app-integrated, web-based, and standalone tools, as well as Artificial Intelligence, into the curriculum. Such modules would equip students and future educators with a deeper understanding of these technologies and how they can be utilized in language education. The implications of these findings are significant, as they provide a roadmap for the development of more effective and engaging AICs for language learning in the future.

Limits and future directions

This study has three main limitations: firstly, the cross-sectional design and reliance on self-reported data may limit the ability to establish causality; secondly, the specific context and non-probabilistic method may restrict the generalizability of findings, necessitating replication in different settings and populations; thirdly, the focus on four specific AICs may not fully capture the complexity of this new technology, underscoring the need to incorporate a broader range of AICs for a more comprehensive evaluation. Further research in the field of AICs could encompass a variety of areas, such as investigating their effect with different language levels, as well as the efficacy of AICs across different student age groups and contexts. Finally, the impact of various chatbot design elements on student interaction and engagement in language learning could be explored. These elements could include multimedia integration, the incorporation of social media, and notably, the use of speech technologies.

Acknowledgements
The authors would like to express their gratitude to all the college students from both institutions for their invaluable participation in this project.

Author contributions
Conceptualization was done by both JB and VK. The methodology was developed by JB. Both JB and VK were responsible for the software and treatment. Validation and formal analysis were carried out by both JB and VK. The original draft was written by JB and reviewed and edited by VK. Both authors have read and agreed to the published version of the manuscript. All authors read and approved the final manuscript.

Funding
Not applicable.

Availability of data and materials
The datasets generated and/or analysed during the current study are not publicly available due privacy reasons but are available from the corresponding author on reasonable request.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 19 September 2023 Accepted: 11 December 2023
Published online: 19 December 2023

References

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications, 2*, 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Ajisoko, P. (2020). The use of Duolingo apps to improve English vocabulary learning. *International Journal of Emerging Technologies in Learning (iJET), 15*(7), 149–155. https://doi.org/10.3991/ijet.v15i07.13229

Annamalai, N., Eltahir, M. E., Zyoud, S. H., Soundrarajan, D., Zakarneh, B., & Al Salhi, N. R. (2023). Exploring English language learning via Chabot: A case study from a self determination theory perspective. *Computers and Education: Artificial Intelligence*. https://doi.org/10.1016/j.caeai.2023.100148

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. https://doi.org/10.1016/j.procs.2015.08.219

Bao, M. (2019). Can home use of speech-enabled artificial intelligence mitigate foreign language anxiety–investigation of a concept. *Arab World English Journal (AWEJ), 5*, 28–40. https://doi.org/10.24093/awej/call5.3

Beattie, A., Edwards, A. P., & Edwards, C. (2020). A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies, 71*(3), 409–427. https://doi.org/10.1080/10510974.2020.1725082

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences, 12*(17), 8427.

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

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*. https://doi.org/10.1080/03055698.2020.1850426

Chuah, K.-M., & Kabilan, M. (2021). Teachers' views on the use of Chatbots to support English language teaching in a mobile environment. *International Journal of Emerging Technologies in Learning (iJET), 16*(20), 223–237. https://doi.org/10.3991/ijet.v16i20.24917

Dokukina, I., & Gumanova, J. (2020). The rise of chatbots–new personal assistants in foreign language learning. *Procedia Computer Science, 169*, 542–546. https://doi.org/10.1016/j.procs.2020.02.212

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

Fryer, L., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology, 24*(2), 8–22.

Fryer, L., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior, 93*, 279–289. https://doi.org/10.1016/j.chb.2018.12.023

Gnewuch, U., Morana, S., Adam, M. T., & Maedche, A. (2022). Opposing effects of response time in human-chatbot interaction: The moderating role of prior experience. *Business & Information Systems Engineering, 64*(6), 773–791. https://doi.org/10.1007/s12599-022-00755-x

Godwin-Jones, R. (2022). Chatbots in language learning: AI systems on the rise. In B. Arnbjörnsdóttir, B. Bédi, L. Bradley, K. Friðriksdóttir, H. Garðarsdóttir, S. Thouësny, & M. J. Whelpton (Eds.), *Intelligent CALL, granular systems and learner data: short papers from EUROCALL* (pp. 124–128). Research-publishing.net. https://doi.org/10.14705/rpnet.2022.61.1446

Gokturk, N. (2017). Mondly learning languages. *Pronunciation in Second Language Learning and Teaching Proceedings, 8*(1), 241–247.

Hajizadeh, S., Salman, A. R., & Ebadi, S. (2023). Evaluating language learning applications from efl learners' perspectives: The case of mondly. *Research Square*. https://doi.org/10.21203/rs.3.rs-3011332/v1

Hakim, R., & Rima, R. (2022). Chatting with AI Chatbots applications to improve English communication skill. *Journal of English Language Studies., 7*(1), 121–130. https://doi.org/10.30870/jels.v7i1.14327

Haristiani, N., Danuwijaya, A. A., Rifai, M. M., & Sarila, H. (2019). Gengobot: A chatbot-based grammar application on mobile instant messaging as language learning medium. *Journal of Engineering Science and Technology, 14*(6), 3158–3173.
Haristiani, N., & Rifa'i, M. M. (2020). Combining chatbot and social media: Enhancing personal learning environment (PLE) in language learning. *Indonesian Journal of Science and Technology, 5*(3), 487–506. https://doi.org/10.17509/ijost.v5i3.28687
Hsu, M.-H., Chen, P.-S., & Yu, C.-S. (2021). Proposing a task-oriented chatbot system for EFL learners speaking practice. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2021.1960864
Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237–257. https://doi.org/10.1111/jcal.12610
Hwang, G.-J., & Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments, 31*(7), 4099–4112. https://doi.org/10.1080/10494820.2021.1952615
Jang, J., Ko, Y., Shin, W. S., & Han, I. (2021). Augmented reality and virtual reality for learning: An examination using an extended technology acceptance model. *IEEE Access, 9*, 6798–6809. https://doi.org/10.1109/ACCESS.2020.3048708
Jeon, J. (2021). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2021.2021241
Jeon, J., Lee, S., & Choi, S. (2023). A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2023.2204343
Jung, S. K. (2019). Introduction to popular mobile chatbot platforms for English learning: Trends and issues. *STEM Journal, 20*(2), 67–90. https://doi.org/10.16875/stem.2019.20.2.67
Khang, A., Muthmainnah, M., Seraj, P. M. I., Al Yakin, A., & Obaid, A. J. (2023). AI-Aided teaching model in education 5.0. In A. Khang, V. Shah, & S. Rani (Eds.), *Handbook of research on AI-based technologies and applications in the Era of the metaverse* (pp. 83–104). IGI Global.
Kharis, M., Schön, S., Hidayat, E., Ardiansyah, R., & Ebner, M. (2022). Mobile Gramabot: Development of a chatbot app for interactive German grammar learning. *International Journal of Emerging Technologies in Learning, 17*(14), 52–63. https://doi.org/10.3991/ijet.v17i14.31323
Kim, N.-Y. (2019). A study on the use of artificial intelligence chatbots for improving English grammar skills. *Journal of Digital Convergence, 17*(8), 37–46. https://doi.org/10.14400/JDC.2019.17.8.037
Kim, Nao-Young. (2020). *Chatbots and language learning: effects of the use of AI chatbots for EFL learning*. Eliva press.
Kim, N.-Y., Cha, Y., & Kim, H.-S. (2019). Future English learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning, 22*(3), 32–53.
Kohnke, L. (2023). L2 learners' perceptions of a chatbot as a potential independent language learning tool. *International Journal of Mobile Learning and Organisation, 17*(1–2), 214–226. https://doi.org/10.1504/IJMLO.2023.10053355
Kukulska-Hulme, A., & Lee, H. (2020). Intelligent assistants in language learning: An analysis of features and limitations. In K.-M. Frederiksen, S. Larsen, L. Bradley, & S. Thouësny (Eds.), *CALL for widening participation: Short papers from EUROCALL* (pp. 172–176). Research-publishing.net.
Li, Y., Chen, C.-Y., Yu, D., Davidson, S., Hou, R., Yuan, X., Tan, Y., Pham, D., & Yu, Z. (2022). *Using Chatbots to Teach Languages. Proceedings of the Ninth ACM Conference on Learning@ Scale*, 451–455. https://doi.org/10.1145/3491140.3528329
Nastasi, B. K., Hitchcock, J., Sarkar, S., Burkholder, G., Varjas, K., & Jayasena, A. (2007). Mixed methods in intervention research: Theory to adaptation. *Journal of Mixed Methods Research, 1*(2), 164–182. https://doi.org/10.1177/1558689806298629
Panesar, K. (2020). Conversational artificial intelligence-demystifying statistical vs linguistic NLP solutions. *Journal of Computer-Assisted Linguistic Research, 4*, 47–79. https://doi.org/10.4995/jclr.2020.12932
Petrović, J., & Jovanović, M. (2021). The role of chatbots in foreign language learning: The present situation and the future outlook. *Artificial Intelligence: Theory and Applications, 2*, 313–330. https://doi.org/10.1007/978-3-030-72711-6_17
Pham, X. L., Pham, T., Nguyen, Q. M., Nguyen, T. H., & Cao, T. T. H. (2018). Chatbot as an intelligent personal assistant for mobile language learning. *Proceedings of the 2018 2nd International Conference on Education and E-Learning*, 16–21. https://doi.org/10.1145/3291078.3291115
Pin-Chuan Lin, M., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot: A mixed-method classroom study. *Journal of Educational Technology & Society, 23*(1), 78–92.
Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies, 151*, 102630. https://doi.org/10.1016/j.ijhcs.2021.102630
Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies, 149*, 102601. https://doi.org/10.1016/j.ijhcs.2021.102601
Vera, F. (2023). Integrating Artificial Intelligence (AI) in the EFL classroom: Benefits and challenges. *Transformar, 4*(2), 66–77.
Yang, J. (2022). Perceptions of preservice teachers on AI chatbots in English education. *International Journal of Internet, Broadcasting and Communication, 14*(1), 44–52. https://doi.org/10.7236/IJIBC.2022.14.1.44
Zhang, R., Zou, D., & Cheng, G. (2023). A review of chatbot-assisted learning: Pedagogical approaches, implementations, factors leading to effectiveness, theories, and future directions. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2023.2202704

Publisher's Note