# CHAPTER 1
# INTRODUCTION

This chapter provides the background and context of the study. It includes a statement of the problem, and rationale of the study. Research objectives, research questions, and the significance of the study are then introduced. Finally, operational definitions of key terms used in this study are defined, with a discussion of the scope and limitations of the study.

## 1.1 Background and Context of the Study

In English language education, four skills - listening, speaking, reading, and writing have been identified of highest importance (Brown, 2000). Among these skills, speaking is the most vital skill to be mastered by learners as it is the manner that language learners measure how much they understand a language, and demonstrate how effectively they communicate with other people (Nazara, 2011). Additionally, the proficient speaking of English can provide students with many potential advantages. One of the advantages is the advancement in language acquisition. Nergis (2020) mentioned that the students with low English proficiency capacity have difficulties focusing on both the form and meaning of the language, which led them to speak slowly with frequent mistakes. Therefore, it can be inferred that if students wish to speak English effectively, they need to improve both fluency and accuracy so that they can speak English with higher ability and less mistakes. Once the students fluency and accuracy are improved, their English speaking skill would likely advance.

Another advantage of proficiency in English is that the students can obtain greater academic achievement. Genesee, Lindholm-Leary, Saunders, and Christian (2006) claimed that the communicative skills and competence allow a student to become successful in academic aspects as proficiency requires the acquisition of vocabulary, grammar, and semantics of a language. It also providers learners with practice actually using the language for communicating with other speakers of that language, from which it is inferred that the student can obtain higher academic success.

A final advantage is that the students can receive better jobs and higher likelihood of promotion of the course of their lifetime. Tuyen and Loan (2019)

reported that the growth of national economy depends on students competence in spoken English in reference to international markets. Therefore, it is predictable that if EFL students are proficient in spoken English, they will receive more chances to have their job of choice and career advancement. As a result, it is valuable to further discuss the current status of how English is taught and learned in Vietnam in general, and how English speaking is important to Vietnamese undergraduate EFL students. It is therefore necessary to find the most appropriate intervention and facilitation to support students' progress.

English teaching in Vietnam has moved forward to communicative-focused instruction as it is an international language. Most of research studies have been conducted to improve communicative competences of the EFL students in English classrooms (Nguyen & Nguyen, 2016). As a result, English speaking becomes an important skill which the EFL students need to achieve (Quyen, Nga, & Nguyen, 2018). Interestingly, the Vietnamese government has driven the country s advancement through nationwide computer investment, and has encouraged teachers to utilize technology in their teaching, to enhance traditional teaching methods. This trend has improved the EFL students positive attitudes toward their studies because they tend to believe that they can be proficient in English when they utilize technologies (Thao et al., 2019).

Within this context, previous studies worked on investigating the effects of using technology in teaching communication skills to the Vietnamese undergraduate EFL students. Duc (2017) applied a computer-based model in helping the students to assess their speaking performance and become more aware of their learning. Quyen and Loi (2018) worked on improving the students English speaking skill by assigning the students to watch and listen to videos spoken by native speakers. Anh and Nhu (2021) employed visual aids to help the students improve English speaking skill by reducing the stress of the speaking environment. All studies mentioned previously reported positive findings and statistically significant differences from their experimental interventions. Therefore, it can be reasonably concluded that using technologies could significantly help Vietnamese undergraduate EFL students improve their English speaking skill.

There are a variety of technologies used for educational purposes such as digital readers and tablets, 3D printing, virtual reality (VR), gamification, cloud technology, artificial intelligence (AI), and mobile technology (Fulton, 2019). Among the technologies for education, artificial intelligence (AI) is one of the most influential technologies and it has been widely popularized in education (Cheng, Sun, & Zarifis,

2020). In Vietnam, AI technologies are suggested for automatic grading of learners exercises, improving their awareness in studies and assessment, helping them to correct grammatical mistakes, and improving the students pronunciation and knowledge of preposition usage.

Nga (2022) suggested that using AI technologies such as ELSA Speak could help students improve pronunciation in speaking. Grammarly could help to fix grammatical mistakes in writing. Turnitin was broadly used by Vietnamese universities to check for students possible plagiarism. Nghi, Phuc, and Thang (2019) reported that using AI technologies like chatbots with conversational activities in the English classrooms could help students learn new grammar knowledge in terms of preposition usage. AI technologies are considered to be one of the effective tools in foreign language education as they can provide flexible, interactive, and learner-center learning which gives them significant potential in EFL contexts, facilitating the enhancement of oral communication (El Shazly, 2020). Studies on use of AI technologies to improve Vietnamese undergraduate EFL students English speaking skill have not been published to date, which inspired the basis for this study.

Having emerged in recent years, AI technologies are used in the fields of robotics, computer vision, speech recognition, and natural language processing (Kaplan, 2016). Nowadays, speech recognition and natural language processing are two main technologies used in developing voice chatbots in order to create the bots which can communicate with a user like a human, called AI voice chatbots. Thanks to the AI technologies, AI voice chatbots can identify questions from an end user and respond with appropriate answers quickly and explicitly based on the assistance of AI algorithms. This technology allows chatbots to be considered a type of e-tutors (Colace et al., 2018).

According to Kim (2016), AI chatbots in general are used as communication tools and are particularly divided into AI text chatbots and AI voice chatbots. AI voice chatbots have been continuously developed with artificial intelligence algorithms and voice recognition technology and can have a voice chat with users, especially EFL students, and help them improve their English speaking skills. Furthermore, Kim (2018) mentioned that the convenience of AI voice chatbots in usage was their rapid and free functions with online access, which allows the students to use them flexibly and freely. Kim, Cha, and Kim's (2019) finding revealed that AI voice chatbots could speak like a native speaker and conduct humanlike conversations with users by with the simulation of an AI-based communication environment, by also imitating human speech patterns, users were convinced that they were talking to a real person.

Based on the findings mentioned previously, it can be predicted that AI voice chatbots may be useful for the Vietnamese undergraduate EFL students because AI voice chatbots were reported to help students concentrate on grammar knowledge, improve English speaking, and provide native-like and humanlike conversations. These findings suggest potential solutions to the improve challenges found in improving Vietnamese undergraduate EFL students English speaking skill.

## 1.2 Statement of the Problems

Previous research studies defined the success of English speaking based on how accurately and fluently a speaker uses the spoken language during a conversation. Actually, Dao (2017) mentioned that the clear communication in a foreign language is achieved when speakers are able to have listeners understand what they mean during a conversation. This achievement requires speakers to be accurate and fluent when they speak the language. Brown (2000) also stated that accuracy and fluency distinctively demonstrate learners language competence. Manurung (2015) further mentioned that fluency and accuracy are the components of English speaking which need to be achieved. Therefore, it can be understood that accuracy and fluency are inseparable elements which determine the success of learners English speaking. However, many researchers on the field of English speaking in Vietnam found that Vietnamese undergraduate EFL students could not speak English successfully due to their insufficient skill, their limited exposure to English speaking environments, and their unwillingness to speak in English classrooms.

Quyen, Nga, and Nguyen (2018) makes reference to the students English insufficiency within the limitations of using correct vocabulary, grammatical structures, and sentences. Therefore, these challenges may hinder the students in their efforts to speak English successfully. Actually, English speaking was found to be the most challenging skill for Vietnamese undergraduate EFL students to achieve successfully, reflected in their low scores for the speaking tests compared to listening, reading, and writing. They could not engage in long discussions without code switching to the Vietnamese language, and they made grammatical and pronunciation mistakes when speaking English (Dao, 2017). This situation was also clarified by Thao and Nguyet (2019) that EFL learners witnessed a high frequency of difficulties in developing language because they lacked grammatical and vocabulary knowledge. In addition, Quyen and Loi (2018) reported that due to the students lack of English proficiency, they did not feel confident speaking in English classrooms. However, it is worth asking why the students had insufficient English speaking

proficiency. To answer for this question, we may argue that apart from the difficulties of language itself, as previously mentioned, the limitation of exposure to English is also likely the primary contribution to the students lack of proficiency.

In fact, this problem has also been reported by some previous studies. Wang (2014) found that English speaking was greatly ignored in EFL classrooms because most of teachers tended to speak a lot in class instead of giving their students chances to speak English. As a result, the students were unable to develop their spoken English, which led to their insufficient speaking competence. In addition, instructional approaches have not always been updated (Chen & Hwang, 2019). The challenge for English speaking is also due to the lack of practice and exposure that the students from non-English speaking countries witness, as they only get English input and practice from language classes, schools, colleges, or universities (Baek & Lee, 2018). Vietnamese undergraduate EFL students also have experienced a similar situation. Quyen and Loi (2018) mentioned that the students in English classrooms were given limited opportunity for practicing English speaking due to insufficient meeting time in English courses. Sharing similar context, Quyen et al. (2018) pointed out the limitation of an English speaking environment caused by the frequent use of the mother tongue in English speaking classrooms and the design of English teaching syllabus and curriculum which did not provide adequate speaking activities for the students. Students need to frequently practice English speaking and see how the language is used in a social setting to become comfortable and fluent (Nguyen & Tran, 2015; Thao & Nguyet, 2019). Therefore, this is another argument that the limitation of English exposure is one of the problems preventing Vietnamese undergraduate EFL students from achieving English speaking skills.

Another problem was found to contribute to the insufficient English speaking skill of Vietnamese undergraduate EFL students, which is the students unwillingness to speak English. According to Tuyen and Loan (2019), the students lack of willingness to speak may cause unsuccessful language production. However, the reasons for the unwillingness of the students in speaking were caused by several factors, among which, the degree of the students motivation was primary. This observation was also reported by Tram (2020), that the students found no motive expressing their ideas in English. Trinh and Pham (2021) further supported this position by confirming that sometimes the students chose not to say anything whenever they felt uncertain with what they were supposed to say, often because they had no ideas about the specific topics they were given. Therefore, we may assume that Vietnamese undergraduate EFL students unwillingness in English

speaking may also relate to their insufficiency in knowledge of the English language, which prevents them from speaking English successfully. This assumption may be right because we know that without learning vocabulary and grammatical structures, Vietnamese undergraduate EFL students may not know how to compose words and sentences to produce the spoken language so that the hearer can understand. As a result, they may be afraid of failure in English speaking, which makes them decide to remain silent.

In summary, it can be implied that the students insufficient English speaking skill, their limited exposure to English speaking environments, and their unwillingness to speak in English classrooms have an intertwined relationship with each other. Vietnamese undergraduate EFL students lack of language proficiency may be due to their limited opportunities for exposure to English speaking, which prevents them from acquiring English speaking skill effectively and cause them be be unwilling or fearful regarding speaking English. Hence, using technologies like virtual reality (VR), gamification, cloud technology, mobile technology, or artificial intelligence voice chatbots (Fulton, 2019) may provide a helpful solution. These technologies provide a new environment for the students to have exposure to English language, which improves their English proficiency and encourages them to learn English more successfully. Although previous research studies found the problems in promoting Vietnamese undergraduate EFL students improvement in English speaking skills, those studies did not consider any forms of technology, such as AI voice chatbots, into consideration for solving the problems. For example, Dao (2017) suggested the involvement of both teachers and students in addressing speaking problems of the students and minimizing the hesitation behind their speaking. Also, Nguyen and Tran (2015) and Quyen et al. (2018) proposed the application of speaking activities, the enrichment of speaking environment for the students to practice English speaking, and the increase in the students awareness of independent learning styles when studying English speaking. Based on the research studies reviewed previously, no technological solutions like AI voice chatbots were suggested to help Vietnamese undergraduate EFL students overcome their difficulties in achieving English speaking skills. Therefore, this study is focused upon finding the effects of an artificial intelligence voice chatbot on English speaking skills of Vietnamese undergraduate EFL students. It will assess the ability to fill the gap between knowledge and experience because AI chatbots were reported to be effective for EFL students in terms of English speaking due to their particular positive features such as learner-center environment, human-like interaction, and authentic input/output (El Shazly, 2020;

Kim, 2018; Kim et al., 2019). Within the scope of this research, the effects of an artificial intelligence voice chatbot on English speaking skill of Vietnamese undergraduate EFL students will be investigated to explore whether an AI voice chatbot can improve the target students speaking skill by enhancing their speaking fluency and accuracy.

## 1.3 Rationale of the Study

Based on the background and context along with the problems mentioned previously, an investigation on the effects of an artificial intelligence voice chatbot on English speaking skill of Vietnamese undergraduate EFL students may provide a unique solution. Therefore, this study is support three key benefits in terms of language exposure, curriculum, and speaking environment. First, an AI voice chatbot will create more chances for the EFL students to become exposed to English speaking. This is important because language exposure, according to Krashen (1982), is the key for learners to achieve higher language proficiency. This is also expected that the EFL students will be able to speak English with more fluency and accuracy when they achieve higher language proficiency. In fact, Kim's (2016) study explored the possibility that chatbots could bring up many opportunities for foreign language learners to practice a target language. Second, this study expects that an AI voice chatbot will be integrated into the teacher s curriculum for teaching English speaking in Vietnamese universities because curriculum and learning environment are among the key factors affecting the students success in English speaking apart from teachers, students, and methods (Dao, 2017). Therefore, the integration of an AI voice chatbot into English speaking s curriculum is predicted to allow Vietnamese undergraduate EFL students to practice English speaking. Then by evaluating themselves, they will have feedback to improve autonomous practice of English speaking. Practically, Fryer and Carpenter (2006) suggested that students could proceed with self-practice and evaluation by doing an exercise of chatting with the chatbots, recording the transcript and then examining it to self-evaluate their speaking performance. Third, a new English speaking environment in association with an AI voice chatbot is hoped to solidly motivate Vietnamese undergraduate EFL students to speak English more frequently and confidently because AI chatbots can be used flexibly in terms of time and place in addition to communicating authentically with the learner (Kim et al., 2019). Within this perspective, Haristiani (2019) reported that AI chatbots motivated students in speaking English because the

students could use the chatbots at anytime, anywhere, and speak more confidently with the chatbots than speaking while facing teachers.

This study will need to gather data to find the significant differences in the English speaking skill between Vietnamese undergraduate EFL students who study in the traditional classrooms compared to those who study by using an AI voice chatbot which was designed for improving the students' English speaking skill. This allows us to investigate the effects of an artificial intelligence voice chatbot on the English speaking skill of Vietnamese undergraduate EFL students to assess potential effects an AI voice chatbot has on Vietnamese undergraduate EFL students in improving English speaking skill. In addition, exploring Vietnamese undergraduate EFL students opinions on using the AI voice chatbots for improving English speaking skill is also an important aspect to assess. Therefore, this study comes up with two research objectives.

## 1.4 Research Objectives

This study has two objectives:

1. To investigate the effects of an AI voice chatbot on the English speaking skills of Vietnamese undergraduate EFL students.

2. To explore Vietnamese undergraduate EFL students opinions on interacting with an AI voice chatbot.

## 1.5 Research Questions

On the basis of the objectives, this study is focused upon the following questions:

1. What are the effects of an AI voice chatbot on English speaking skill acquisition of Vietnamese undergraduate EFL students?

2. What are the opinions of Vietnamese undergraduate EFL students about using an AI voice chatbot?

## 1.6 Significance of the Study

This study is conducted to investigate the effects of an AI voice chatbot on the English speaking skills of Vietnamese undergraduate EFL students. Therefore, it is expected that with its findings, this study will enhance learning experiences, provide for pedagogical improvement, and add to the research data on this topic.

Regarding learning enhancement, this study is expected to contribute an option for students to motivate them to speak English in classroom environment as Kim (2016) reported that the AI chatbots helped to enhance the students positive attitude to motivate the students to speak in a relaxed atmosphere. Also, the introduction of an AI voice chatbot in this study will provide learners with more learning options. In fact, Nghi, Phuc, and Thang (2019) found that when class meetings are combined with the chatbots, they elevate students performance when they learn the lessons of the course book. In addition, through the investigation of Vietnamese undergraduate EFL students opinions about using an AI voice chatbot to improve English speaking skills, this study may propose vital adjustments for Vietnamese undergraduate EFL students to enhance their ability to self-practice. Technological skills may also improve, because Thao et al. (2019) found that the integration of technology-enhanced language learning (TELL) tools into an English language learning (ELL) curriculum can increase the students autonomy and centeredness.

For pedagogical enhancement, teachers may benefit from using an AI voice chatbot as a teaching assistant for teaching English speaking because Colace et al. (2018) mentioned that AI voice chatbots can play the role of e-tutors thanks to the association of natural language processing (NLP) algorithms and the ontology model which can help chatbots understand user swords to successfully communicate with users. Moreover, teachers may also need to integrate an AI voice chatbot into their English speaking curriculum to enrich their instructional resources. El Shazly (2020) suggested that the integration of AI applications into English classrooms can optimize learning experiences.

For research data, researchers on technology-enhanced language learning (TELL) may benefit from this study as it will explore and report possible effects of an AI voice chatbot on English speaking skill of Vietnamese undergraduate EFL students. The findings in this study may become useful clues for future researchers to continue investigating further effects of an AI voice chatbot on enhancing receptive skills and other productive skills of Vietnamese undergraduate EFL students when they learn English. This perspective is also supported with the point of view from Kim et al. (2019) that the effects of the use of AI voice chatbots should be shifted to the investigation of the four basic language skills: listening, reading, speaking, and writing.

## 1.7 Scope of the Study

Based on the significance previously mentioned, this study is narrowed to the investigation of the effects of an AI voice chatbot on English speaking skill of Vietnamese undergraduate EFL students. Previous studies proved statistical significance even though they were conducted with small and medium-scale sample sizes (Han, 2020; Kim, 2016, 2017, 2018; Kim et al., 2019; Nghi et al., 2019). This study will conduct its research with an AI voice chatbot only, as AI voice chatbots in general were reported to be able to communicate through not only text messages but also voice messages (Han, 2020). For the effects of an AI voice chatbot on English speaking skill of Vietnamese undergraduate EFL students, this study will mainly delve into investigating the effects of an AI voice chatbot on Vietnamese undergraduate EFL students English speaking skills in terms of fluency and accuracy when they speak English, because these two components were reported by previous researchers to be the central tasks to the success of EFL learners in English speaking.

## 1.8 Operational Definitions of Key Terms

**The artificial intelligence voice chatbot** or **the AI voice chatbot** in this study is a program integrated with artificial intelligence algorithms and can talk with users through the text-to-speech and voice recognition processes operating under internet connection. According to Kaplan (2016) and Colace et al. (2018), an AI voice chatbot is a kind of program or app which works under artificial intelligence algorithms such as machine learning combined with natural language processing operated with internet connection and can have a chat with users via a mobile device. In this study, Andy English Bot, called Andy, is an AI voice chatbot used thorough the experiment as it was reported by Kim et al. (2019) that it could chat with users in form of text chat and voice chat and can help users learn and practice English conversation, vocabulary, and grammar.

**The Vietnamese undergraduate EFL students** in the study means the non-English majored undergraduate students whose English level is equivalent to A2 level in CEFR standard, classified by a TOEIC Bridge test. Those students are in the first year or second year and learn English 2 which is one of the three modules of General English taught by Faculty of Foreign Languages at Can Tho University, Vietnam. They come from various schools and colleges in Can Tho University such as College of Natural Sciences, School of Education, College of Technology, College of Agriculture.. Accordingly, Vietnamese undergraduate EFL students are re ed by those

who major in non-English academic programs such as Chemistry Education, Agriculture, Primary Education, and Computer Networking.

**Traditional classrooms** mentioned in this study refers to the English classrooms in Can Tho University where students study on textbooks and follow teachers guidance in the courses. Traditional classrooms are conducted in a face-to-face manner and English language skills are taught by a teacher using oral instruction, handouts, or electronic media such as CDs, MP3 files, or presentation slides throughout the lesson.

**AI-aided classrooms** suggested as a term in this study implies the English classrooms in Can Tho University in which an AI voice chatbot is integrated into the teaching curriculum. The AI-aided classrooms are also conducted in a face-to-face manner but the students learn English speaking through a website and practice English speaking with the integrated AI voice chatbot while the teacher only provides technological support to the students.

**English speaking skill** in this study refers to the students competence of demonstrating accurate and fluent spoken language to communicate during a conversation in English as a foreign language, from which the hearer can understand the speaker s message. This key term is based on Dao (2017) mentioning that English speaking skill is achieved when the speaker can successfully convey his meaning to the listener during a conversation by speaking the language fluently and accurately.

**Fluency** is defined in this study as one of the distinctive criteria demonstrating learners proficient speaking skill. Fluency is the flowing or the natural language produced by the EFL students when they speak English, demonstrated through speaker s appropriate hesitations and pauses. Brown and Lee (2015) explained that fluency in speaking is best achieved when the speech flow is performed smoothly. In addition, Wang (2014) determined hesitations and pauses in speaking as the main obstructions to speaking fluency.

**Accuracy** in this study is defined as the EFL students correct use of pronunciation, vocabulary, and grammar when they speak English. According to Brown and Lee (2015), accuracy focuses on the accurate use of phonology, grammar, morphosyntax, and discourse in the spoken output. Wang (2014) also stated that speaking accuracy involves the correct use of pronunciation, vocabulary, and grammar in English speaking. Together with fluency accuracy is seen in this study as another distinctive criteria to demonstrate learners' proficiency in speaking skills.

**Opinions** mentioned in this study refers to the Vietnamese undergraduate students opinions on using the AI voice chatbot for learning English speaking skill,

classified into four themes: practice process, fluency, accuracy, and feeling. This term was composed by this study to address four main aspects of the investigation referring to (1) how long and how frequently the students practiced English speaking with the AI voice chatbot, (2) to what degree the students improved their English speaking fluency, (3) in what manner their English speaking accuracy was improved, and (4) how the students felt after using the AI voice chatbot.

## 1.9 Summary

After introducing the purpose and scope of the study, including the intending instruments and outcome measures, the next topic is a critical review of the literature including previous empirical studies related to this study topic.

# CHAPTER 2
# LITERATURE REVIEW

The literature review firstly delves into English speaking skill to explore how fluency and accuracy are closely intertwined with each other in evaluating the success of learners' English speaking skill based on previous research study's findings. Secondly, a review on the potential use of technology to improve English speaking skill will be discussed. Thirdly, a thorough review of AI chatbots including definitions, characteristics, operational structures, and types of AI chatbots will be discussed. This is followed by the review on AI voice chatbot applications in the EFL contexts to discover their potentials in enhancing English language skills in EFL contexts, in particular, in learning English speaking as a foreign language. Next, a review of how Vietnamese undergraduate EFL students learn English speaking in traditional classrooms will be conducted to discover the current standard of the learning process which may have caused difficulties for the students achieving English speaking skills successfully. This will allow introduction of this research which will propose an additional method to help the target students overcome the difficulties. Finally, this chapter will conduct a literature review on technology-enhanced language learning (TELL) which holds the theoretical framework of this research.

## 2.1 English Speaking Skill

Brown and Lee (2015) determined that reading, listening, writing, and speaking were the four most important skills for EFL learners to achieve in English language education. Among the skills, Nazara (2011) stated that speaking is the most essential skill required to be developed by EFL students in order to communicate effectively. Speaking skill is a skill set of the students regarding producing words, language, or expressing ideas (Socheath, 2018). Therefore, speaking is determined as a productive skill in a language. According to Brown and Lee (2015), learners should learn speaking skill with a message-oriented approach in which the focus is on speakers ability to communicate in different contexts so that listeners can understand. Brown (2000) stated that accuracy and fluency are important goals to pursue in spoken language, and they determine the success of English speaking. Hedge (2000) proposed accuracy

and fluency to be the criteria for evaluating speaking. Accordingly, an L2 speaker is determined to be a competent L2 speaker if he or she at first can use English accurately in terms of pronunciation, grammar, lexis, with appropriate use in specific contexts. Second, he or she should speak English fluently. Manurung (2015) suggested that successful speaking skill is aimed at the mastering of speaking fluency, accuracy, and comprehensibility. According to Brown and Lee (2015), learners fluency and accuracy demonstrate correct comprehension in English speaking and infer language acquisition. Therefore, it can be understood that observing the learners fluency and accuracy in English speaking can reflect learners English speaking skill. However, accuracy and fluency are two connected aspects because learners require both for successful speaking skills. Actually, Brown (2000) stated that when learners pay too much attention to fluency in speaking, they try to speak quickly and smoothly, and then they tend to use wrong grammar, pronunciation, and inappropriate vocabulary and vice versa. In other words, the faster they try to speak, the more mistakes are found during their speech. Therefore, it is a recommendation that teachers focus their teaching approach to a message-oriented mode so that learners can initially use the language, and then progress to using it properly.

Previous researchers have different views on speaking accuracy and fluency. Wang (2014) determined accuracy in English speaking as the correct use of pronunciation, vocabulary, and grammar. Accuracy from Brown and Lee s (2015) point of view, is the accurate use of voice and articulation, grammar, and phonology. Dao (2017) related speaking accuracy with the learners correct use of grammar and pronunciation. According to Socheath (2018), speaking skill requires learners to have the knowledge of vocabulary, parts of speech, pronunciation, expression, sentence structures, and tense. Therefore, this suggestion implies speaking accuracy in terms of vocabulary, pronunciation, and grammar (elements of speech, pronunciation, expression, sentence structures, and tense).

Among the research studies reviewed, the work of Wang (2014) and Socheath (2018) provides a more comprehensive discussion of speaking accuracy when compared with other research studies reviewed. Other authors, such as Brown (2015) and Dao (2017) relate speaking accuracy with only the correct use of grammar and pronunciation. Fluent English speaking is explained by Walker and White (2013) as the ability of L2 speakers to maintain the smooth flow of the speech without hesitations and in a coherent manner, from which they will maintain the thought flow. Within this perspective, it can be seen that the main purpose of speaking is to

negotiate meaning of the messages between the speaker and the hearer of the language, while linguistic features are to support the process of meaning negotiation.

This author concludes that accuracy and fluency can be considered the main principles of successful English speaking skills. The next section will provide a review of the potential use of technology to improve students English speaking skill.

## 2.2 The Potential Use of Technology to Improve English Speaking Skill

In our daily life, we tend to communicate with each other through the applications on the internet such as Zoom, LINE, Facebook Messenger, and Telegram or through text messages. In those cases, we use written language to communicate. For example, we tend to write or type hey", ur", lol", or 555" in our messages to simulate speaking and expressing our emotions. Regarding this phenomena, Walker and White (2013) stated that there were no clues of spoken language found in such types of communication as we could not see any body language or intonation to confirm it was a real speech, but the use of written language could still provide a variety of opportunities fostering learners ability to practice a second language such as the speech transcribed from movies and plays. Along with written language, spoken language could be found in the communication between people through webinar platforms such as Skype, Zoom, Google Meet or in the communication between humans and the AI chatbot applications such as Google Diagflow, Andy English Bot, Siri, ELIZA, Duolingo, or SpeechAce.

These examples are the specific to platform and situation. Through the webinar platform, we can see everyone s body language and hear their voice and intonation. For the AI chatbot applications, although the visuality of body language cannot be observed, we can still hear the AI speaker s voice and intonation. Walker and White (2013) suggested that rehearsing spoken language before having a conversation in real life is a way to help students familiarize themselves with real life speaking situations. This suggestion requires AI voice chatbots which can help with this process when they can perform repeatable question-answer processes (Kim, 2017). They can speak with students about the same topic tirelessly and attentively (Fryer and Carpenter, 2006). It is necessary to delve into a discussion about AI voice chatbots and their applications for learning English in EFL contexts and to provide details of how this technology can help with enhancing EFL students language proficiency.

## 2.3 Artificial Intelligence (AI) Chatbots

The term artificial intelligence (AI) chatbots" is well known to researchers in this area. Fryer and Carpenter (2006) reported that chatbots have witnessed a long history of their development since the first appearing version, called ELIZA in the early 1960s. In defining an AI chatbot, Haristiani (2019) stated that an AI chatbot is a computer program which is able to carry out conversations through audio or text. When it is integrated with artificial intelligence, it has the ability to update the knowledge from previous conversations. Kim et al. (2019) also describe AI chatbots as agents which engage in a conversation through providing their users with the natural language interface to simulate a human-like interaction with the users if integrated with artificial intelligence. Adamopoulou and Moussiades (2020) related AI chatbots with artificial conversation entities (i.e. the agents which are able to interact with users, the intelligent bots, such as digital assistants such as Siri, Cortana, and Alexa).

### 2.3.1 Characteristics of AI Chatbots

Previous studies determined the characteristics of AI chatbots in terms of repeatability, flexibility, productivity, and innovation. With repeatability, Fryer and Carpenter (2006) confirmed that an AI chatbot could discuss the same material with students repeatedly and endlessly as it did not feel bored or tired as human being might. Moreover, when speaking with an AI voice chatbot, the students could often repeat or rephrase an utterance to negotiate meaning (Kim, 2017). Regarding flexibility, an AI chatbot can perform both text and speech conversations with the students (Fryer and Carpenter, 2006). Importantly, an AI chatbot is available on the internet and can be downloaded onto smartphones at no cost (Kim, 2017; Kim, 2018). Ahmad, Che, and Zainal (2018) also mentioned that an AI chatbot can receive and respond to human inquiry 24 hours a day. In terms of productivity, Ahmad et al. (2018) and Adamopoulou and Moussiades (2020) mentioned that the best benefit of an AI chatbot is that it can serve a great number of users at the same time and reach a broad audience on a messenger system. Finally, an AI chatbot is innovative because it is continuously updated with artificial intelligence and voice recognition technology (Kim, 2016).

Moreover, Colace et al. (2018) discussed using an AI chatbot as the most innovative solution erasing the distance between technology and education as it could provide the students with interactive learning experiences, enhance skills, and provide a personalized learning environment. Kim (2018) concluded that an AI chatbot could perform human-like English conversation with the students as native speakers, giving the students authentic language input. Based on the characteristics

described, many studies were conducted to investigate the potentials of AI chatbots on language learning and found that AI chatbots could help with students communication skills, provide the students with negotiation of meaning, and increase the students motivation and interest in learning. In the next section, this study reviews the operational structures of AI chatbots to explore how they are constructed and operated.

**2.3.2 Operational Structures of AI Chatbots**

Previous researchers described the structure of AI chatbots in different ways. However, the components used for operating an AI chatbot were described quite similarly from one researcher to another; and the AI chatbot in their descriptions seems to serve the same purpose. Different operational structures of AI chatbots proposed by previous studies are reviewed below.

Ahmad et al. (2018) conducted a study to investigate the structure of AI chatbots in terms of their working system and structuring techniques. The researchers used Chat.io, Collect.chat, and Cleverbot as the samples for their investigation. The results showed that the AI chatbots worked under a chain of text processing which runs in the background and are supported by a variety of techniques such as Artificial Intelligence Markup Language (AIML), Pattern Matching, Language Tricks, Chatscript, Parsing, Database, and Markov Chain. It is worth examining two specific parts which are the text processing and the techniques associated in the process. Figure 2.1 shows the chatbot system processes adopted from Ahmad et al. s (2018) s work.

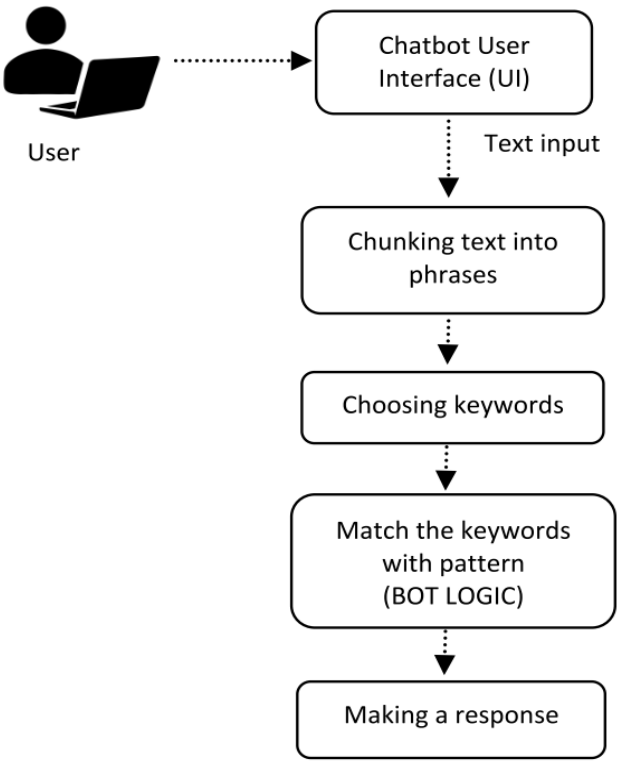



**Figure 2.1 Chatbot System Processes (Ahmad et al., 2018, p. 9)**

The text processing starts when a user inputs text into the user interface. At that time, the system will receive and analyze the user s text by separating it into small words and phrases to treat those items as the keywords and then matching them with the text pattern stored in the database before processing the output in form of text response.

With regard of the techniques used by an AI chatbot, the researchers regarded the AIML as a core unit which administers the process of modeling the conversation into a response process, followed by other supportive techniques. First, the Pattern Matching generates appropriate responses from matching the user s questions with the existing conversational patterns. Second, the Learning Tricks enriches the AI chatbot s knowledge base with additional words, phrases, or sentences to provide more persuasive responses to a user. Third, the Chatscript, including the written scripts coded by developers, provides syntax elements for the system so that an AI chatbot can give plausible responses even if it fails in matching the user s text with the existing conversational pattern. Fourth, the Parsing uses natural language and computer language to analyze whether a text contains any semantic or further information. This technique refers to the Natural Language Processing which was defined by Liddy (2001) as the computerized approach used to

analyze texts based on both a set of theories and a set of technologies. Fifth, the Markov Chain technique builds the consequential responses in order to make better sense. Finally, the database stores the conversational data including previous keywords and patterns so that an AI chatbot can remember and perform more successful matches when giving responses to a user. One thing to be noticed in this study is that AI chatbots at that time could communicate with users via text input and output.

Colace et al. (2018) administered a study investigating an AI chatbot system designed for supporting educational purposes through the question-answer communicating process. Pandora, Forensic Challenger (TFC), IBM Bluemix, and Facebook Messenger Bot GUI were taken as samples of the AI chatbot system which were used as a case study and experiment to study an operational framework for designing an AI chatbot. The results of the study showed that when an AI chatbot was integrated with the Natural Language Processing (NLP) algorithms and an ontologies domain, it could detect user s messages and respond appropriately. When compared with the work of Ahmad et al. (2018), the writers of this study found that Colace et al. s (2018) work described further development of an AI chatbot design when he engaged conversational flow into the evaluation to perform successful conversation and handle user requests. Conversational flow is defined by Koudenburg, Postmes, and Gordijn (2014) as *the extent to which a conversation is experienced as smooth, efficient and mutually engaging."* The reason why the conversational flow is seen to be a further step toward the development of an AI chatbot relates to the complexity of its structure. In Colace et al. s (2018) system architecture shown in Figure 2.2, an AI chatbot did not work under the simple text processing based on the techniques described by Ahmad et al. (2018) any longer. Instead, it worked under the conversational flow design which engaged four main operational items: Front-End, Back-Office, Knowledge Base Module, and E-Learning BOT Module.

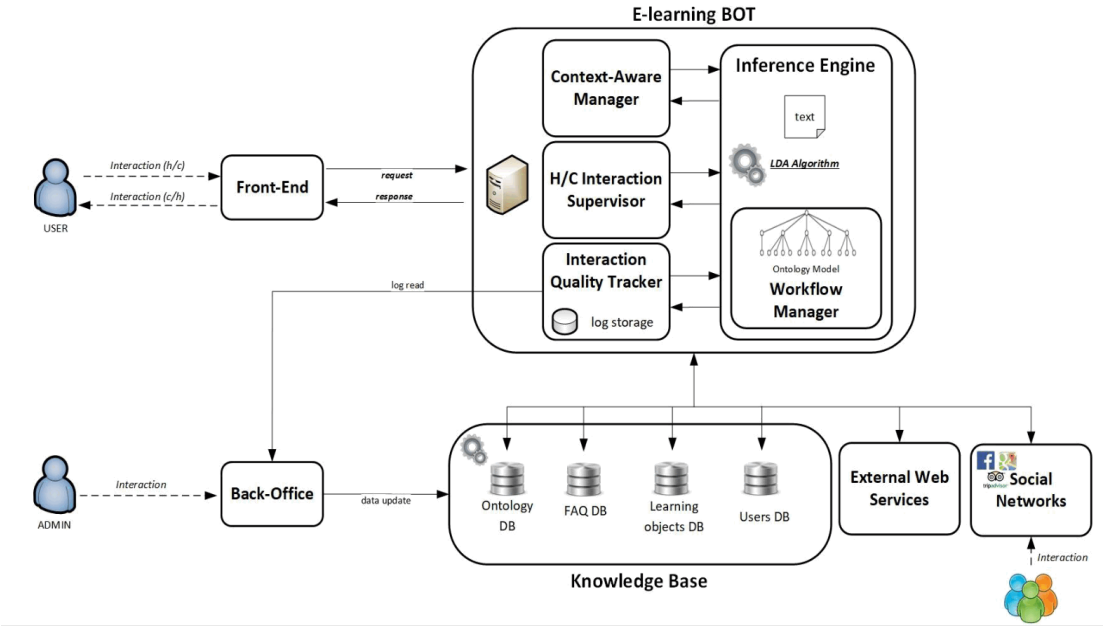



**Figure 2.2 System Architecture *(Colace et al., 2018, p. 531)***

The Front-End is to provide a user-friendly interface. The Back-Office works in the background to manage the operations and is invisible to users. The Knowledge Base Module manages the database and processes the data. Finally, the E-Learn BOT Module was the core engine containing four smaller units: Interaction Quality Tracker, Human-Computer Integration Supervisor, Context-Aware Information Manager, and Inference Engine. Specifically, the Interaction Quality Tracker unit is used to monitor interactions between an AI chatbot and its users. The Human-Computer Integration Supervisor unit supervises dialogue, identifies ambiguous questions, and analyses the need for community support in case the system failed to answer the questions. The Context-Aware Information Manager unit drives dialogue based on context; and finally, the Inference Engine unit releases accurate responses to the user through the Latent Dirichlet Allocation Algorithm (LDA) and the database of the Workflow Manager. The LDA and the Workflow Manager are the important parts of an AI chatbot as they can manage and adopt the definition of an ontology or of pre-existing ontologies. They man then look up the ontology through word analysis based on the conversation to select appropriate sentences for an AI chatbot to provide correct answers to user s questions.

As previously mentioned, Colace et al. s (2018) work involved two important items which helped an AI chatbot realize a user s question to respond correctly. Those items are the Natural Language Processing algorithms and the ontologies domain. Therefore, a further step is needed to examine Colace et al. s

(2018) model of ontology and reference framework to see how the Natural Language Processing algorithms and ontologies domain work within the AI chatbot system. In Colace et al. s (2018) Ontology Model shown in Figure 2.3, the ontology model is built up under five elements: Topic of Study, User, Course, Lesson, and Learning Object.

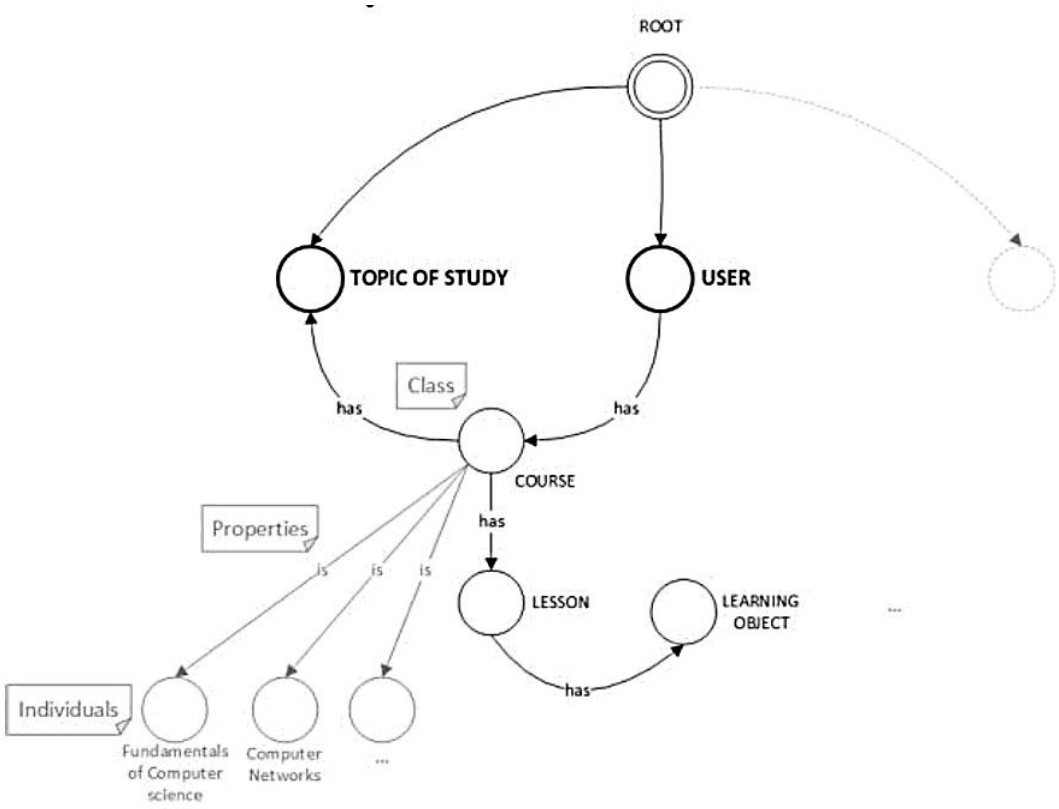



**Figure 2.3. The Ontology Model *(Colace et al., 2018, p. 531)***

The first element refers to the knowledge of a set of subjects belonging to specific fields of study such as English Language Studies, Computer Sciences, or the like, reflected by the second element which is the User. The User contains both students and teachers who contribute their knowledge of a specific field of study into the ontology system. Such the knowledge is constructed from the remaining elements which are the Course (the smaller part or the topic of the study field such as English speaking, English listening, or the like), the Lesson (the specific module of the course), and the Learning Object (the digital or web-based resource supporting the learning).

However, how to allow an AI chatbot to utilize the ontologies domain to learn" remains a question. In this case, the answer relies on the operation of the

Natural Language Processing (NLP) algorithms demonstrated in Colace et al. s (2018) reference framework shown in Figure 2.4.

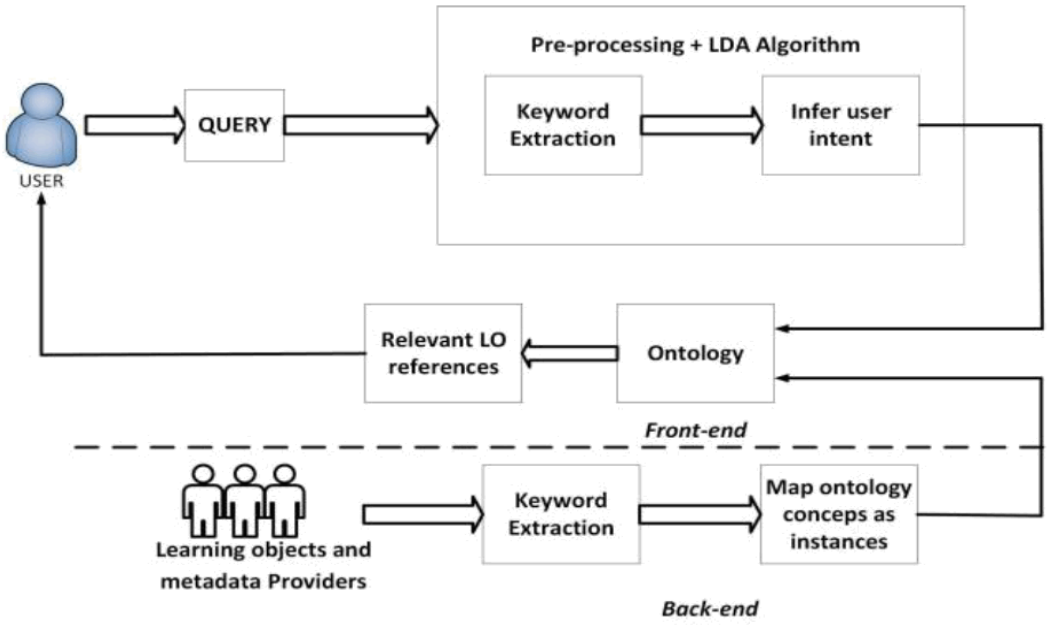



**Figure 2.4 The Reference Framework *(Colace et al., 2018, p. 531)***

Within this framework, the NLP algorithms fulfill three tasks. The first task is to build an e-learning ontology which has the capacity of addressing the relationship between learning objects, evaluating the semantic distance between them, and then inferring from a user s intention to clarify the user s question. The second task is to manage the questions asked by users to serve the purpose of extracting keywords to thereby infer user intention. This task is done by a semantic inference engine administered by the Latent Dirichlet Allocation (LDA) algorithm which helps an AI chatbot connect the questions made by users with the existing learning object data to perform successful semantic inference. The final task is to map the instances of the existing learning object into an ontology so that an AI chatbot can have concepts or examples to infer the relations between learning objects. One thing to be aware of is that AI chatbots at that time were still unable to perform oral communication with the user. However, the text-to-speech technology was considered to be a promising factor which might be integrated into an AI chatbot so that it could respond to the user in both text and voice forms.

Adamopoulou and Moussiades (2020) conducted a study about the history, technology, and applications of an AI chatbot to perform an in-depth investigation on its development, technological design, and applications in various

industries. The researchers took a wide range of AI chatbot applications into the investigation such as MSN ELIZA, Apple Siri, IBM Watson, Google Now, Google Assistant, Microsoft Cortana, Microsoft XiaoIce, and Amazon Alexa. They further classified AI chatbots into specific categories: knowledge, service, goals, response generation, human-aid, permissions, and communication. Adamopoulou and Moussiades (2020) clarified that a communicative AI chatbot is used as a communication channel which involves text, voice, and image into the communication and interaction processes with their user. The findings of the research shed light on the AI chatbot s techniques and architecture. If compared with the work of Ahmad et al. (2018) and Colace et al. (2018), it can be assumed that Adamopoulou and Moussiades (2020) work on AI chatbots has been a crucial development because it includes nearly all aspects of the other contributors while adding more innovative approaches in terms of pattern matching and machine learning techniques.

In Adamopoulou and Moussiades s (2020) model, the pattern matching technique involves three programming approaches which include Artificial Intelligence Markup Language (AIML), RiveScript, and ChatScript. The AIML is an open-source language which was developed in 1995, using XML (eXtensible Markup Language) for implementing codes for the purpose of creating natural language within dialogues to facilitate the communication between human and AI chatbots (Adamopoulou & Moussiades, 2020; Marietto et al., 2013). XML is easy for designing and implementing the codes thanks to its tag structure. For the example shown in Figure 2.5, if we wish the system to start text processing when a user types Hi!", then the tag structure will be <pattern>Hi!</pattern>. After that, with the attribute named <template>...</template> as the leading tag, developers can list all possible random discourses such as Hi! How are you?", Hi there!", or just Hi!" into subordinate tag <random>...</random> from which the bot can randomly select the listed discourses to respond to the user. Within this structure, <template> or <random> is an opening tag, </template> or </random> is a closing tag, and the "..." stands for the information or discourse to be used by an AI chatbot. The reason why tag structure is preferred by developers is because of the simple structure of opening tag < > and closing tag </ >. Developers only need to remember the correct name of the commands to put into the tags and then the system will carry out the rest process. ALICE and ELIZA were the predecessors of AI chatbots which were integrated with the AIML language.

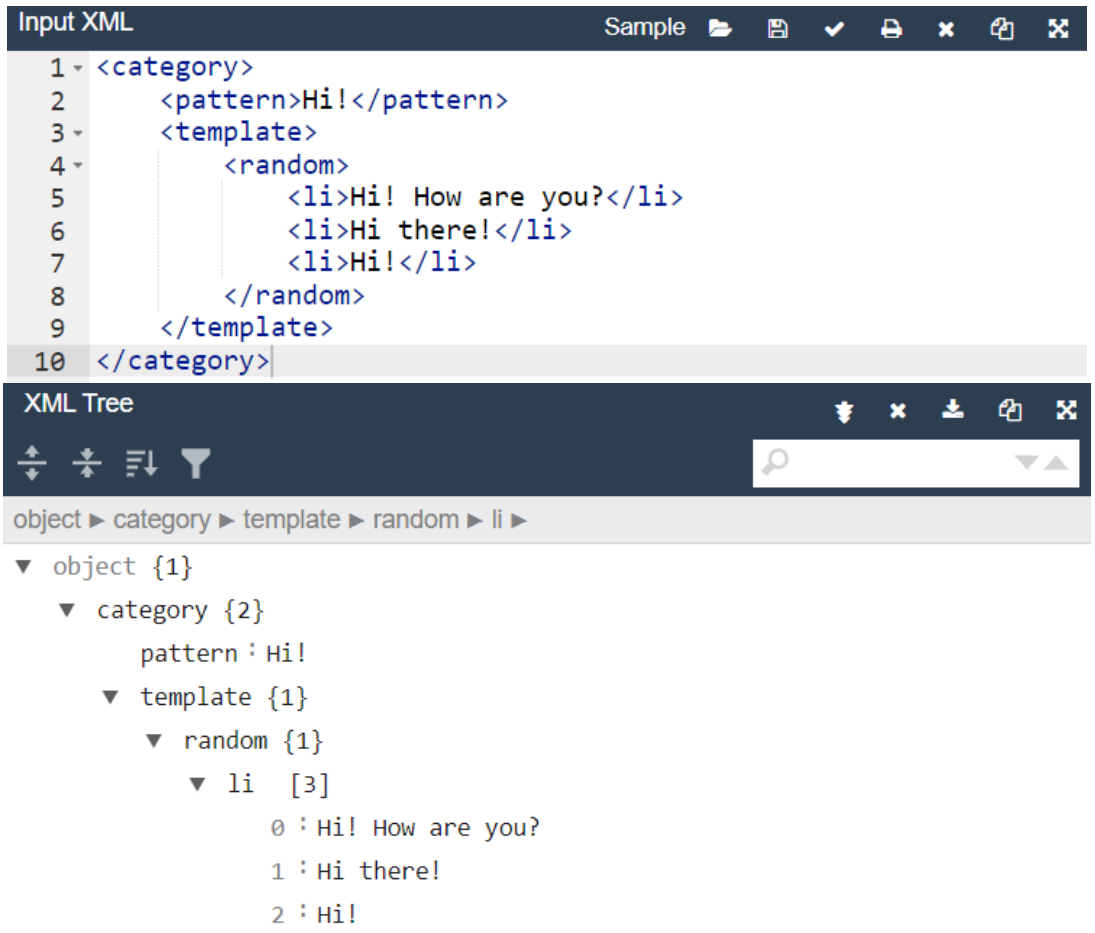



**Figure 2.5 The eXtensible Markup Language (XML)**

If compared the AIML approach with the other two, we may find that the RiveScript and ChatScript are more innovative but complex. This should not be surprising as the AIML language was created more than twenty years ago while the RiveScript and ChatScript namely were invented in 2009 and 2011. Moreover, the AIML is an open source language working under XML environment while the RiveScript operates under Java and Python environments, and ChatScript utilizes a completely new programming language. What is similar between the three approaches is that all of them are open-source languages which can be modified or coded by any developer or coder. Adamopoulou and Moussiades (2020) described the RiveScript as *a line-based scripting language implementing the Knowledge Base in rule-based chatbots"*. In the Java and Python coding environments, the RiveScript appears in the form of +", -", and *". The symbol +" means the user s input and the -" refers to an AI chatbot s responses. Finally, the *" is to mark important conversations where the input of the interpreter unit operated by the Natural Language Processing algorithms in an AI chatbot s Knowledge Base is required. To

provide a further explanation of RiveScript, this study will adopt the sample script of Adamopoulou and Moussiades (2020) shown in Figure 2.6 and give a detailed explanation about the script.

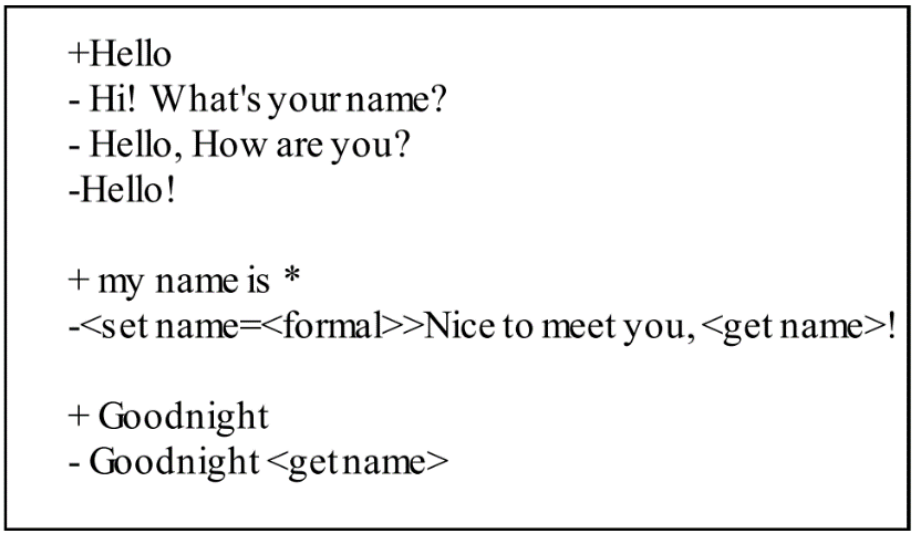



**Figure 2.6 Example of RiveScript code** ***(Adamopoulou & Moussiades, 2020, p. 6)***

We may separate the conversation into three phases. The first phase starts from +Hello" and ends with -Hello!" The second phase comprises two next lines, and the final phase is the balance. In the first phase, the code structure +Hello" means that, when a user types Hello", an AI chatbot may respond with either Hi! What s your name?" or Hello, How are you?" or just Hello!". In the second phase, the *" infers that whenever a user types their name in that space, another process will be carried out by the interpreter unit in the Knowledge Base. This process will be implemented by Java or Python coding language, read by the formula <set name=<formal>>...<get name>!> in which the "..." stands for the discourse to be uttered by an AI chatbot. Within this aspect, the conversation in the second phase will start when a user introduces his name (i.e. My name is Peter). At that time, the formula mentioned above will inform an AI chatbot that the name of the user is 'Peter,' and that this type of introduction is a formal type. Therefore, the suitable response should be Nice to meet you, Peter!". In the final phase, when a user types Goodnight", then the same process will take place as in the first and the second phases so that the AI chatbot can respond appropriately by saying Goodnight Peter."

The ChatScript technique as mentioned previously is the latest programming language for operating an AI chatbot. It is also an open-source

programming language as RiveScript but it is much more complicated as it further involves the embedding of tagger and parser analyses (Adamopoulou & Moussiades, 2020) into the process. This allows an AI chatbot to further analyze the linguistic features and semantic meanings from the input to enable a high-proficient response. Dramatically, ChatScript was reported to be able to handle both short-term memory processing and long-term memory processing to help the AI chatbot memorize the user s information from previous conversations for later use. Unfortunately, this programming language was reported to be only suitable for the developers who already possessed a high level of expertise in programming languages due to its complexity.

Regarding learning machine techniques, the AI chatbot models by Adamopoulou and Moussiades (2020) support dramatic development because they operate with a wide range of AI approaches including the Natural Language Processing (NLP), the Natural Language Understanding (NLU), the Artificial Neural Networks, the Recurrent Neural Networks, the Sequence-to-Sequence model, and the Deep Seq2seq Models . In contrast, Ahmad et al. (2018) and Colace et al. s (2018) models only utilized the Natural Language Processing system. The Natural Language Understanding algorithms work on an AI chatbot s response generation process which serves to support the Frequently Ask Question functions. Accordingly, the NLP and the NLU are significant approaches of machine learning. In fact, if the NLP is to help the system to translate and monitor natural language, then the NLU helps the system to understand the natural language, and therefore, the NLU can be seen as an inseparable companion of the NLP. The Natural Language Understanding (NLU) is the algorithm which helps with the process of classifying the intents and extracting the entities in accordance within a context. At this point, there are two things we need to consider: the intents and the entities. The intents are seen as specific topics which may be derived from the prediction of the users messages. Figure 2.7 provides an example for an intent which has been developed for this study, powered by Google Diagflow, which is a platform for creating an AI chatbot.

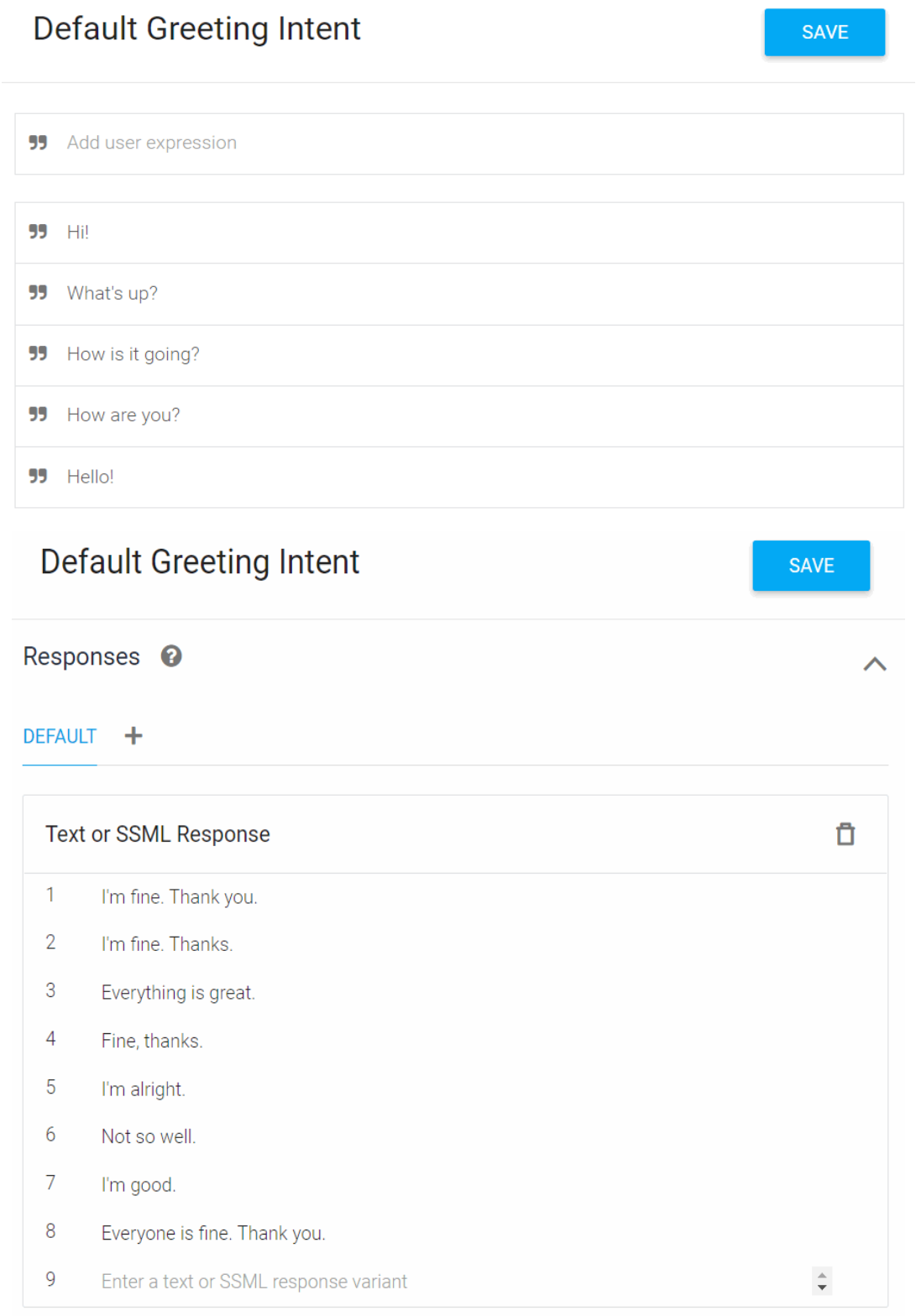
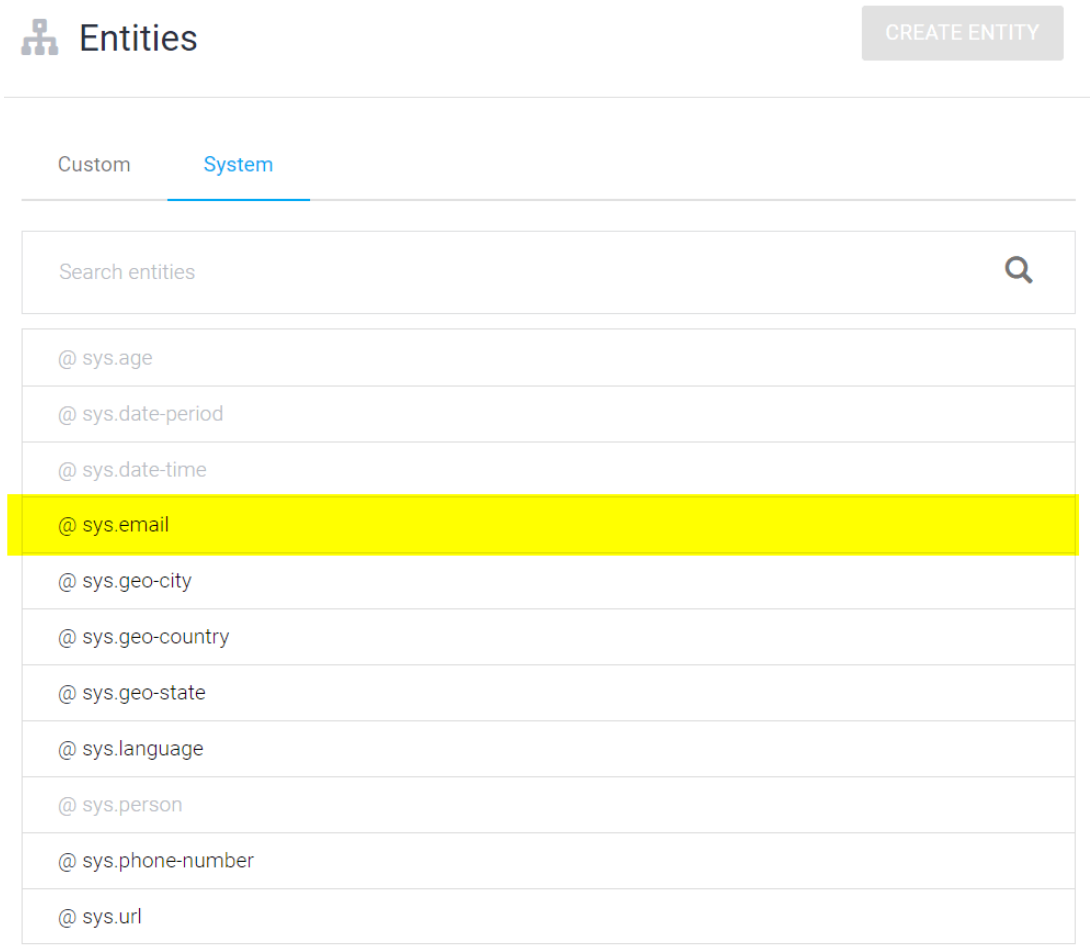



**Figure 2.7 Example of An Intent**

Within this example, we can predict that when performing greetings, one can use a variety of discourses to say hello!" such as hello!", hi!", how are you?", how are you doing?", or the like. Within this situation, an intent can be created and labelled as Greeting" in which the possible discourses related to greeting will be stored in the intent together with possible responses that an AI chatbot can utilize to respond to the user.

The entities are seen as the word blocks which contain the definition within a context and can be defined manually by developers or automatically by a given system. Figure 2.8 gives an example for an entity generated under Google Diagflow platform.



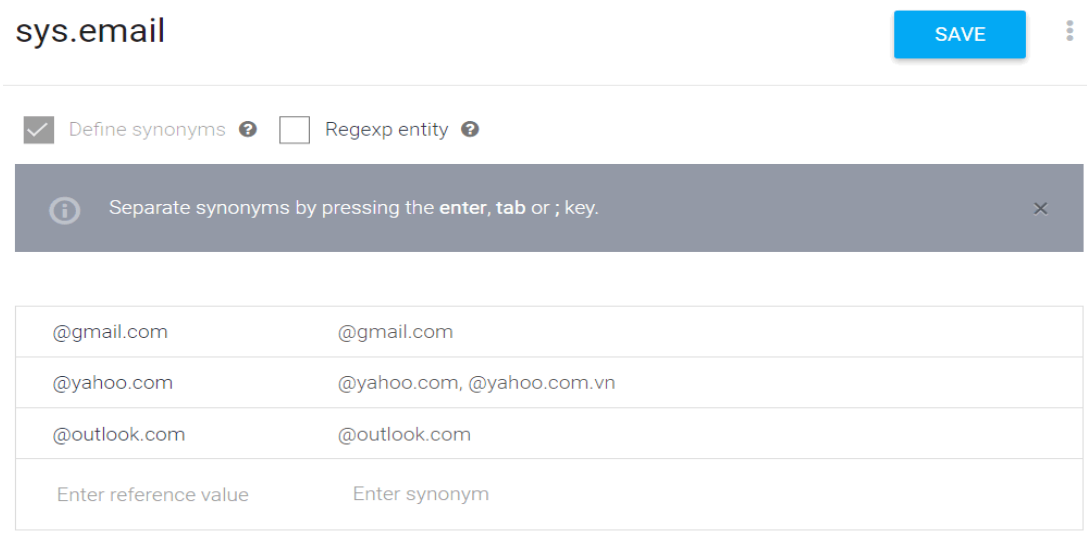



**Figure 2.8 Example of Entities**

This example is used to explain that when a user provides his email address during the conversation with an AI chatbot, the email addresses may come with the formats of @gmail.com, @yahoo.com, @outlook.com, or the like, and therefore each email format may have millions of characters put before the @" letter because there are millions email users who use these various email services provided those providers. The users must use a different email address from each other. The entities, in this case, will help an AI chatbot to recognize and understand that email addresses which contain the formats mentioned above are email addresses, and produce an appropriate response rather than become absorbed in detailed processing for meaning.

The last but important aspect to be considered in Adamopoulou and Moussiades s (2020) AI chatbot model is the system structure which contains five elements: user interface, user message analysis, dialogue management, the backend, and response generation. Figure 2.9 illustrates the system structure of an AI chatbot in Adamopoulou and Moussiades s (2020) model.

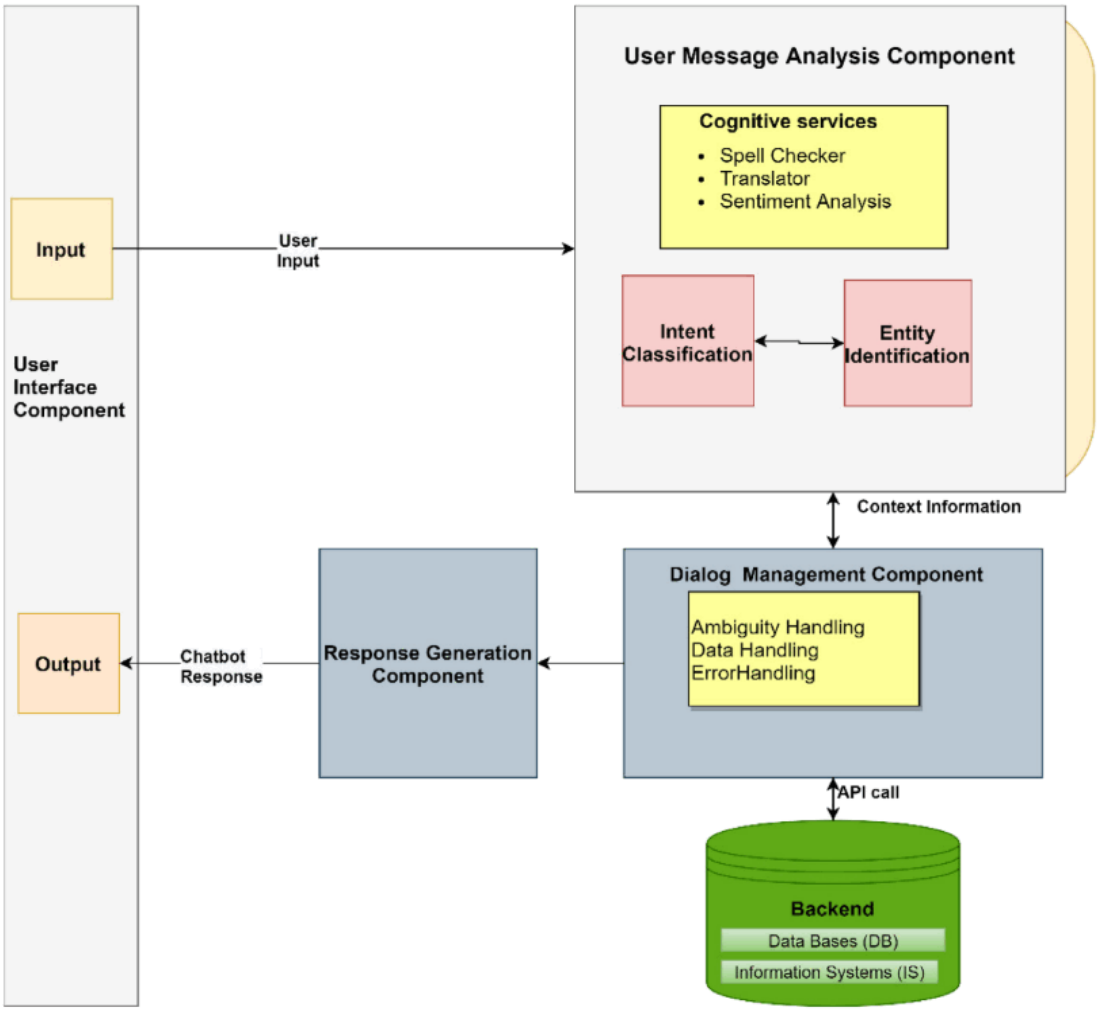



**Figure 2.9 General Chatbot Structure** ***(Adamopoulou & Moussiades, 2020, p. 11)***

Basically, this AI chatbot structure shares some similar features as Colace et al. s (2018) model in terms of its Front-End design. The Back-End design where many tasks are processed in the background seems to be somehow different. For the Front-End, both models have the same structure of User Input and Response Output which is responsible for receiving user s messages and producing responses to those messages in a visible way. For the Back-End, Adamopoulou and Moussiades s (2020) model has four main components. They are the User Message Analysis, the Dialog Management, the Backend, and Response Generation. First, the User Message Analysis is similar to the E-Learning Bot component found in Colace et al. s (2018) AI chatbot structure because this component does the task of message analysis based on the Natural Language Processing (NLP) algorithms. In detail, when a user types a message through the user interface, that message will be sent to the User Message Analysis Component in which the NLP algorithms will analyze the message by

classifying it with the intents and then extracting the entities in accordance with the built-in pattern. This allows the AI chatbot to make inferences from the message before responding to it. A new item found in Adamopoulou and Moussiades s (2020) model is that apart from NLP, the User Message Analysis component also utilizes the Natural Language Understanding to provide an additional function for an AI chatbot, called the Cognitive Services. With the NLU, the Cognitive Services can perform further tasks such as checking spelling, translating, and semantic analyzing. This additional function has not been found in either Ahmad et al. (2018) or Colace et al. s (2018) models. Second, the Dialog Management is responsible for managing and updating the context of a conversation through three modules: Ambiguity Handling, Data Handling, and Error Handling. These modules work by retaining previous conversational information, updating new conversational intents, and retrieving identified entities to suggest necessary discourses for an AI chatbot to inquire for a user s clarification whenever it fails to understand the context of the conversation.

This component is quite similar to Colace et al. s (2018) E-Learning BOT component which is also responsible for handling the issues of conversational contexts through three units: Context-Aware Manager, H/C Interaction Supervisor, and Interaction Quality Checker. Third, the Backend is where the knowledge or ontologies domain is stored in a database or in information systems so that an AI chatbot can select suitable information to answer the user s queries. Colace et al. s (2018) model also has this component with the same function, called the Knowledge Base. Finally, the Response Generation generates the appropriate responses for an AI chatbot, retrieved from the Backend and analyzed by the Dialog Management. This component is quite similar to Colace et al. s (2018) Inference Engine which also works on creating suitable responses to user s messages. One difference between the two components is that Colace et al. s (2018) model used the Latent Dirichlet Allocation (LDA) algorithm to perform the task while Adamopoulou and Moussiades s (2020) model utilized the Natural Language Generation (NLG) algorithm.

Having carefully investigated three AI chatbot structures, it is reasonable to conclude that AI chatbots have witnessed significant development in recent years as they feature innovative techniques such as pattern matching and machine learning. In addition, the AI chatbot has been developed with more functionalities, driven by both conventional and modern programming languages such as AIML, RivesScript, or ChatScript, operated under innovative coding environments such as XML, Java, or Python, and processed by the AI algorithms such as the Natural Language Processing (NLP), the Natural Language Understanding (NLU), the Latent Dirichlet Allocation

(LDA), or the Natural Language Generation (NLG) which serve both pattern matching and machine learning techniques. Moreover, among the three structures analyzed, we find that the Adamopoulou and Moussiades s (2020) AI chatbot model is preferred as it refers to most of the innovative technologies we have just mentioned even though it has some complex programming languages which are only suitable for senior developers. Moreover, programming languages may not be crucial for the researchers on AI chatbots because there are ready-made platforms which can handle all the coding and programming tasks. Such may include Google Diagflow, Dasa, or NativeChat. Within these platforms, researchers need not compose the scripts anymore. Instead, they provide the discourses (also known as parameters) into the given sections, which include intents and identified entities. Then those platforms will automatically conduct the rest of the process. The next section will discuss the difference between an AI text chatbot and an AI voice chatbot.

### 2.3.3 The Difference Between an AI Text Chatbots and an AI Voice Chatbots

Basically, AI voice chatbots have similar operational structures to those discussed previously in this chapter. The only difference between the two types is that an AI voice chatbot is integrated with voice recognition and text-to-speech algorithms which help the system to recognize words and sentences from spoken speech instead of text entry. Some AI chatbots are classified as AI text chatbots because they only support conversations with users in form of texts. The well-known examples include ELIZA, ALICE, and NativeChat. Apart from those, a variety of AI chatbots are considered to be AI voice chatbots due to their capability of communicating with users by both voice and text. The popular names of these are Cleverbot, Andy English Bot, Replika, Apple Siri, Google Assistant, and Amazon Alexa. An interesting point which differentiates an AI text chatbot from an AI voice chatbot is that the latter has the function of speaking, illustrated by the symbol of a speaker or a microphone next to the chat line in the user interface, as illustrated in Figure 2.10.

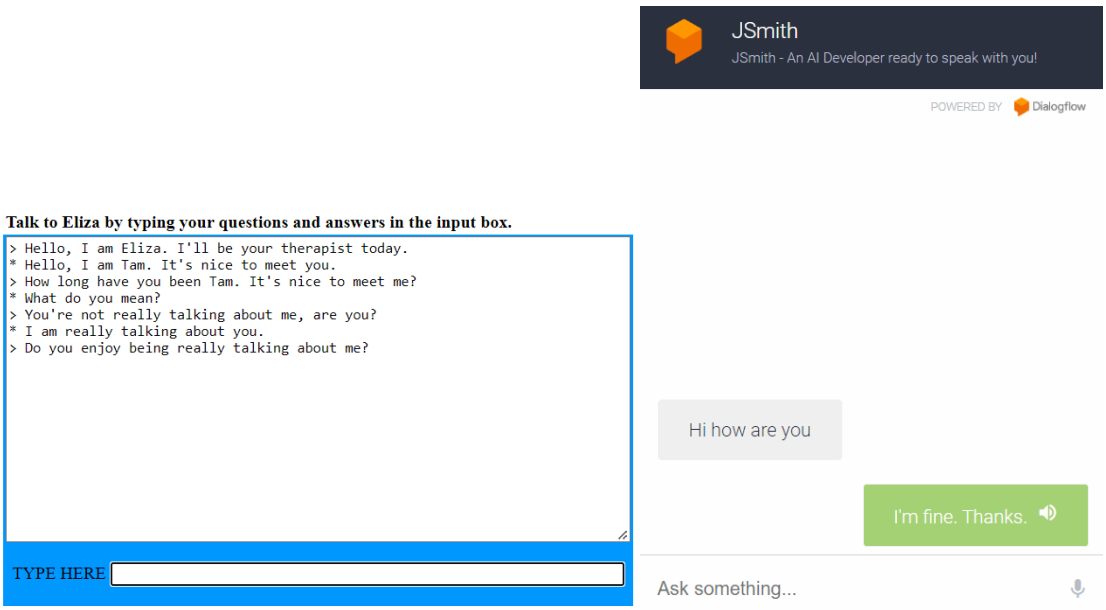



**Figure 2.10 An AI Text Chatbot (left) and An AI Voice Chatbot (right)**

Based on the functions of the two types of AI chatbots, it can be assumed that an AI voice chatbot is more innovative for virtual communications as it can communicate with users through both text and voice messages. Therefore, it should be considered for integration into online learning. The next section will discuss possible applications of various AI voice chatbots in the context of EFL, and students perceptions, to investigate the most suitable AI voice chatbot to employ in this study.

## 2.4 Previous Studies about AI Voice Chatbots in EFL Contexts

AI voice chatbots have become common for customer assistant services. For example, Google Assistant and Apple Siri are two customer service assistants which can assist Android or iOS users. They are able to perform simple tasks like opening an app, finding a song or a piece of online news through voice commands. Amazon Alexa with voice assistance is also integrated into Amazon s products so that it can support customers to perform actions such as turning on a product or switching it off, playing a song or moving forward to another song, as examples. However, less is known about the role of an AI voice chatbot in the educational context, especially in EFL. This section will explore the research studies related to EFL.

Kim et al. (2019) conducted a study reviewing different types of AI chatbots used for language learning. A variety of AI chatbots were accounted in the study such as ELIZA, ALICE, Cleverbot, Elbot, Talk to Eve, Replika, Lyra, Andy English Bot, Poket

Friend, Mondly, and Duolingual. Within the study, the researchers employed nine criteria to evaluate the AI chatbots. Those criteria were (1) the ability of chatbots to understand complex user input, (2) the turn-taking scheme, (3) the ability to recall user s name, (4) the capability of multilanguage support, (5) the feature of voice input and output, (6) the feature of text input and output, (7) the ability to access historical conversations, (8) the ability to deal with strange questions, and (9) the ability to overcome typographical errors. The findings revealed that most of the AI chatbots could perform turn-taking and feature text I/O while some of them could support voice I/O, multilanguage, and recalling historical conversations. Few of the AI chatbots were able to perform other tasks like understanding complex user s input, recalling user s name, answering complicated questions, or overcoming typographical errors. Kim et al. s (2019) AI chatbots data when related to the identified functionalities, found that ALICE, Replika, and Andy English Bot contained many more features than the other programs, especially the feature of voice input and output. However, it is curious that Kim et al. s (2019) finding identified ALICE as the AI chatbot which supported voice input and output because no researchers have mentioned this feature of the program. To test this statement, we attempted to verify the claim by having a chat with ALICE and found that it could not perform any voice input or output. In Figure 2.11, even though it has a button labeled  Say", the button had no functional capacity. The only thing happened after pressing  Say" was that the text we input would appear in the chat session. Therefore, Replika and Andy English Bot remain the primary AI voice chatbots.

Nghi et al. (2019) implemented an experimental study on applying an AI chatbot for helping students to learn prepositions. The study involved 200 undergraduate students divided into two groups: 100 students in the control group and 100 in the experimental group. The researchers of the study utilized the Facebook chatbot developed by Chatfuel as a research instrument to conduct the experiment. The English Pronouns and Prepositions book was used as an instructional instrument during the experiment period. The control group took 15 periods to learn and the experimental group took 10 periods. The results of the study provided three useful pieces of information. First, the AI chatbot brought a new learning experience to the students. The new learning experience was the combination of both the class meeting and the use of the AI chatbot. This combination was observed to accelerate the students  performance in learning. Second, the students  perception of using AI chatbot tools for the learning process was increased because they felt comfortable to self-practice and self-access in learning with the AI chatbot. Third, the AI chatbot

created a fun and exciting environment for the students to learn, which motivated them to share their understandings and performance with classmates.

El Shazly (2020) conducted a research study investigating the role of AI chatbots in English speaking practice for foreign language anxiety (FLA) management of the Egyptian EFL students. The research invited 48 undergraduate Korean EFL students to take part in the eight-week quasi-experimental research. Pre-speaking and post-speaking tests were developed to evaluate the students speaking proficiency level before and after the intervention of AI chatbots. The IELTS speaking test rubric was used and developed for the speaking evaluation. The FLA questionnaire containing 33 items was used to explore the students anxiety levels at pre-intervention and post-intervention. Mondly and some web chatbots such as Audrey, Charles, Cristal, and Mike were used as AI chatbot tools for the intervention. The results of the research revealed that the students speaking proficiency level improved after the intervention. However, it was found that the students experienced a degree of anxiety and that their anxiety increased slightly even after the pre-intervention of the AI chatbots. Therefore, this finding was reported to contrast with other research studies as those always found positive effects of AI chatbots toward EFL students anxiety level. Nevertheless, El Shazly (2020) also explained for this contradictory finding could possibly be explained. First, eight-week intervention was a short duration. Therefore, the students might not have enough time to get familiar with the AI chatbots. Second, the emotional deficit of the AI chatbots might fall short of learners emotional needs, while a real teacher could do so. Third, unfamiliar and irrelevant topics along with repeated conversations with the virtual partners might also demotivate the students, even though the students did improve their language proficiency. Disregarding those negative factors, El Shazly (2020) reported that the AI chatbots could help increasing the students effort. The chatbots were also reported to be able to enhance the students oral performance by improving their cognitive faculties and linguistic abilities.

Kim et al. (2021) conducted research aiming at exploring the effects of voice chat on the communication skills of the EFL learners. The research involved 49 undergraduate EFL students in two different English proficiency levels who took an experiment with fourteen weeks of sessions speaking in voice chat sessions with the AI voice chatbots named respectively Replika, Andy, and Google Assistant. The participants could choose to talk with any of the given AI voice chatbots they preferred. The researcher used questionnaires, pre-speaking and post-speaking, tests, and interviews as the main instruments for data collection procedure. The results of

the research revealed that all students improved their pronunciation, intonation, and stress after the intervention. However, insignificant differences were found on the low-level students fluency improvement while the students at intermediate level could significantly improve their speaking fluency. In exploring the students perceptions on communicating with the AI voice chatbots. Kim et al. (2021) found both positive and negative perceptions from the students. In light of the positive perceptions, the study reported that they could find more opportunities to practice speaking without the fear of losing face or worrying about using the wrong sentences. They could also improve pronunciation and become more confident, active, and interested in speaking due to the unlimited time available when speaking with the AI voice chatbots. Regarding negative perceptions, the study found that the students were not quite comfortable with having voice chats with the AI voice chatbots. Moreover, the study also reported the AI voice chatbots limitations in voice recognition when the students voices were not recognized correctly by the AI voice chatbots. The study also mentioned that the AI voice chatbots could not help the students with English correction.

In investigating the effect of an AI voice chatbot interaction on EFL students L2 speaking performance and anxiety, Çakmak (2022) selected 90 EFL students from a Turkish university and let them use Replika for practicing English speaking during a twelve week period. A questionnaire was utilized to explore the students perceptions of using the AI voice chatbot for practicing English speaking. The results of the study revealed both positive and negative findings. For the positive findings, Çakmak (2022) found that using Replika to practice English speaking could help enhance the students English speaking performance more effectively than having them practice speaking face-to-face with peers. However, the research reported that there was not a significant difference with the students anxiety reduction because of challenges in getting the AI voice chatbot to understand them correctly, which then increased their anxiety more when speaking. Therefore, Çakmak (2022) suggested that using chatbot interaction proved not to be an effective way to help the EFL students reduce anxiety with speaking English.

## 2.5 How Vietnamese Undergraduate EFL Students Learn English Speaking in Traditional Classrooms

Learning English speaking skills in Vietnamese universities is performed in traditional classrooms in which students rely on teachers and textbooks to learn English speaking. However, this traditional learning style creates some difficulties for

Vietnamese undergraduate EFL students to achieve successful English speaking skills. To seek solutions, this study will explore the context of learning English speaking skills in traditional English speaking classrooms and the issues faced by Vietnamese Undergraduate EFL Students.

**2.5.1 Traditional English Speaking Classrooms in Vietnam**

Duyen and Huan (2017) determined the elements of traditional English speaking classrooms by identifying four elements which included context, teacher, student, and tasks. Within this framework, English speaking classrooms were often conducted in terms of group work containing three to five students to allow them to develop communication proficiency in English. However, in such classrooms, the teacher was seen as the one who had authority over students throughout the students learning process, causing the students to passively learn. Tasks and detailed instructions were given and guided by the teacher in which students were instructed to do the group work. Sufficient time was provided for the students to complete the tasks. Although in the group work, communication between students was crucial, they tended not to communicate with each other in terms of asking questions, giving comments, responding with gesture like nodding heads, or by using sounds to show their agreement or disagreement to other members points of view. In these classrooms, the students rarely used English to communicate or shared ideas with their classmates.

Dao (2017) described the traditional English speaking classrooms in terms of material, method, environment, teacher, and student. The study reported that in traditional classrooms the students learned English speaking with the textbook as the only material and very few communicative activities were organized in such classrooms. Furthermore, CD or audio files spoken by native speakers were often used by teachers as the speaking models for students to imitate and the speaking exercises taken from textbooks were given to the students to practice English speaking in front of the class. In such classrooms, the students studied in large-size classrooms accommodating around thirty to forty seats, and teachers used both English and Vietnamese to teach English speaking skills or give instructions to ensure the students understanding of the lesson and tasks. The students used Vietnamese in traditional English speaking classrooms when discussing or debating in group. Regarding the use of Vietnamese in English speaking classrooms, Anh (2010) confirmed that using the Vietnamese language in English speaking classrooms was part of the teaching method. This could be positive in some specific contexts iwhen teachers could decide when to use Vietnamese.

Nghi (2019) reported that in traditional classrooms, language teachers normally taught a foreign language lesson with four stages including warm-up, pre-practice, control practice and post-practice. During classroom time, the students were provided with handouts, teachers gave lectures throughout the lesson, and the classroom time was spent strictly on the course procedure. Thao and Nguyet (2019) reported that at the end of each English speaking course, the students were required to take an individual verbal test with lecturers by having a face-to-face conversation with their lecturers on a chosen topic.

In observation of the traditional English speaking classrooms, Nghi (2019) reported that the normal and repeated procedure from one lesson to another caused boredom for the students. Tuyen and Loan (2019) added that in such classrooms, most of the students never or rarely volunteered to speak or answer the teacher s questions, and often avoided the teacher asking additional questions. As a result, the students often only sat silently in the class and listened to the teacher. Trinh and Pham (2021) reported that in traditional classrooms, the students tended to translate the information in the textbook into Vietnamese before they did the tasks assigned by teacher.

In consideration of the types of traditional English speaking classrooms, a clearer picture of how Vietnamese undergraduate EFL students learn English speaking has been revealed. The traditional classrooms were mainly driven by teachers. Textbooks, groupwork tasks, handouts, and media files were used as the main tools for the teaching and learning process. Moreover, the students still did not feel comfortable speaking in the class and the teacher still had to provide all information during class time. Moreover, the use of the Vietnamese language in those English speaking classrooms was found to happen regularly. Consequently, these limitations might challenge EFL students who planned to achieve successful English speaking skills.

### 2.5.2 Difficulties of Vietnamese Undergraduate EFL Students in Achieving English Speaking Skill in the Traditional Classrooms

In investigating the difficulties of Vietnamese undergraduate EFL students in achieving speaking English successfully, previous researchers found that the students low proficiency in English, their unwillingness to speak, and the lack of English speaking environment were the primary problems in the traditional classroom model.

Dao (2017) conducted a study to investigate the key factors which had negative effects on Vietnamese undergraduate EFL students. The study had 108 non-

English majored students at a university in Vietnam. A questionnaire and classroom observation were employed as the research instruments for data collection. The findings revealed that the lack of English speaking environment, the students insufficient English language competence, and their unwillingness to speak English caused difficulties for students who wished to achieve English speaking skill.

In more detail, the research found that the frequent use of Vietnamese in English classrooms created the habit of using Vietnamese in verbal interactions among the students, which then caused a lack of exposure to English even though they were studying in an English classroom. Regarding the students themselves, the study found that the students were reluctant to speak English due to their insufficiency of vocabulary and grammar to express their thoughts. Furthermore, due to lack of practice in speaking English, the students were then further unwilling to speak English in the class. In addition, factors such as inadequate materials and activities for practicing English speaking, the insufficient time the students were allowed to practice English, and the limitation of space for students to practice speaking English when communicating with teachers and with each other were all found to limit the students exposure and success in speaking English.

As a result, those students failed to realize if they had spoken English successfully or if there were areas needed for improvement. As a solution for the problems found, the study suggested engaging both teachers and students in overcoming the difficulties. Specifically, recommendations for teachers were to speak English in the class more frequently and encourage the students to exclusively use English rather than their mother tongue language in classroom discussion. In addition, the study also mentioned the need for the students to actively speak English in class to form the habit of speaking English among themselves. Moreover, the students were encouraged to actively join English clubs and societies outside the classrooms to increase chances to use English for speaking in the real-life. In line with this study, Nguyen and Nguyen (2016) suggested raising the students awareness of using communication strategies by integrating those strategies into the teaching curriculum to help the students successfully overcome the difficulties of speaking English.

Similarly, Quyen et al. (2018) carried out a study aiming to investigate the factors contributing to the difficulties Vietnamese undergraduate EFL students have been facing when they spoke English. There were 131 students participating in the research: a questionnaire, individual interviews, and class observation were used as the research instruments. The findings revealed that the lack of vocabulary,

confidence, strategies, chances, and motivation contributed to the students difficulties in English speaking.

Furthermore, the factors such as teachers methods, teaching curriculum, in-class environment, and limited extra-curricular activities were found to have created a challenge for the students English speaking skills. Among those factors, the lack of in-class environment and extra-curricular activities outweighed the others because most of the students agreed that they lacked native lecturers to practice English speaking with. They also mentioned that English use was limited after the class. In short, they lacked environments to practice English speaking skill and became fluent.

Some solutions to overcome these difficulties included suggestions that teachers should use various teaching strategies in flexible manner to get the students to use less mother tongue based language in the class so that they would have more opportunity to speak English. In addition, the self-awareness of the students regarding independent learning was also mentioned as one of the possible solutions for the problem.

Duc (2017) further suggested the use of a computer-based model to help the students self-practice English speaking skills. The research used a computer-based model for the pre-speaking and post-speaking tests as the research instruments to investigate if the students were able to self-practice and self-assess English speaking skill after the intervention. There were 40 undergraduate students joining the research which lasted five months. Among the students, 21 were in the experimental group and got familiarized with computer-based speaking tests while the remaining 19 students were in the control group and were not trained to be familiarized with computer-based speaking tests.

The results of the study revealed that using the computer-based model could effectively help the students with self-practice and self-assessment of their speaking and thereby become responsible for learning English speaking skills.

An inference can be drawn that using technology can help with improving Vietnamese undergraduate EFL students English speaking skill by giving them more chances to be more self aware in learning English speaking skill.

Tuyen and Loan (2019) conducted a study in a university in Vietnam to investigate the EFL students willingness to communicate in English and the factors contributing to this. The research chose 200 EFL students and used questionnaire surveys, semi-structured interviews, and class observation as the instruments for the research. The results revealed that most of the students were not willing to

communicate in English speaking classes and their unwillingness was resulted from their insufficiency in English language and shortage of confidence. Furthermore, the research provided an explanation that the students low confidence in English speaking was because the students felt that their English speaking was not proficient enough, and mistakes might lead to the teachers misunderstanding or might make them feel embarrassed.

In addition, difficult tasks and unfamiliar topics were found to impede the students from speaking due to their low proficiency in grammar and vocabulary. To improve the students willingness to communicate in English classrooms, the research suggested improving syllabus design, correct placement of students at the correct level of English skills to help them feel more comfortable.

Also, changing teaching strategies were recommended to help improving the students willingness to communicate in English classes. For this aspect, Anh and Nhu (2021) also stated that visual aids can help speakers release their stress when speaking by shifting the audience s attention from the speaker to the video aids. Therefore, it is understandable that visual aids may also be a useful application of technology to help the students become more willing to communicate in English classrooms as they feel more relaxed in this situation.

Thao and Nguyet (2019) implemented research on examining the four aspects of speaking difficulties faced by the EFL students in a university in Vietnam. The research involved 150 participants who were undergraduates. It employed a questionnaire as the instrument for investigation. Throughout the data collection procedure, the research found that the students had difficulties speaking English although they had already spent considerable time in practice. Importantly, the difficulties were found to be triggered because the students did not have sufficient chances to practice English speaking when they left the classrooms, and they found it difficult to cooperate with their classmates. In addition, the duration of English learning was reported to impair improvement if the time for practice was too short. Long English learning times might allow the students to explore helpful strategies to overcome their difficulties in English speaking. The researchers recommended the students use of recording software on a computer to help them self-create opportunities for practicing English speaking and then having a native speaker evaluate the voice.

The research studies mentioned above suggest that it can be observed that although the students lack of English speaking exposure, their insufficient English language proficiency, and their unwillingness to communicate prevented

them from achieving English speaking skills successfully, there are, nevertheless, possible interventions to help them overcome these problems.

Apart from applying traditional methods such as the application of intensive speaking in English classrooms and the encouragement of frequent speaking in English after class (Dao, 2017), the improvement of students self-awareness (Nguyen & Nguyen, 2016), or the use of speaking strategies (Quyen et al., 2018), recent researchers have gradually shifted their attention to the use of technology to help solving the students' difficulties in English speaking.

Some of the technological applications can be found in previous research such as the use of videos to help the students prepare for speaking in English classrooms (Quyen & Loi, 2018), the application of computer-based models to improve the students autonomy in practicing English speaking skill (Duc, 2017), the use of visual aids to reduce the students stress when speaking in English classrooms (Anh & Nhu, 2021), or the use of computer software to evaluate the students English speaking skill and provide feedback to them for improvement in English speaking skills (Thao & Nguyet, 2019).

However, none of the studies have ever considered the use of an AI voice chatbot as a possible solution. Therefore, it is a reasonable consideration to conduct research on an AI voice chatbot to investigate its effects on the English speaking skills of Vietnamese undergraduate EFL students. Technology-enhanced language learning (TELL) provides the theoretical guideline for using AI voice chatbots in learning English speaking process, which is the next area of discussion.

## 2.6 Technology-Enhanced Language Learning (TELL)

Technology-enhanced language learning (TELL) fosters the use of technology in the English learning and teaching process which has been applied widely in institutions to facilitate the learning objectives of students (Devlin, Feldhaus, & Bentrem, 2013). Theoretically, TELL is a wider range of computer-assisted language learning (CALL) engaging not only computers into the language learning process but also various types of technological devices such as phones, game consoles, and tablets (Walker & White, 2013).

### 2.6.1 The TELL Model

In exploration of the framework applied for the research, the TELL model proposed by Walker and White (2013) will be discussed in detail, illustrated in Table 2.1.

Table 2.1 TELL Model *(Walker & White, 2013)*

| **Approach** | **Structural/ restricted CALL** | **Communicative CALL Open CALL** | **Integrative CALL** | **TELL** |
|---|---|---|---|---|
| **Technology** | From mainframe to mobile | PCs | Multimedia, internet | Mobile devices, tablets, multiplayer games, virtual worlds |
| **English-teaching paradigm** | Grammar-translation and **audio-lingual** | Communicative language teaching | Content-based ESP/EAP | Communication, interaction |
| **View of language** | Structural (a formal structural system) | Cognitive (a mentally constructed system) | Socio-cognitive (developed in social interaction) | Structural, cognitive, socio-cognitive, adaptable |
| **Principal use of technology** | Drill and practice | Communicative exercises | Authentic discourse | Normalized |
| **Principal objective** | Accuracy | Fluency | Agency | Autonomy within community |
| **View of learning** | Behaviourism | Constructivism | Social constructivism/ situated learning | Connectivism |
| **Role of technology** | Tutor | Tutee | Mediational tool | Environment, resource |

In this model, there are seven approaches throughout the development stages from restricted CALL to TELL. In Restricted CALL approach, CALL is determined under the behaviorist view in which the teaching method focuses on grammar translation and audio-lingual data. The language learning comprises both a formal and structural system. The model by Walker and White (2013) seems to further engage accuracy to be the objective of language use and technology plays the role of a tutor who teaches learners. This can be implied that within the Restricted CALL approach, technology has already played a significant role during the complete teaching process.

In the Open CALL approach, the main technological device is the personal computer (PC) and communicative language learning is the teaching method. Within this stage, the language is supposed to deal with not only structures but also cognitive factors. There is a significant development within this approach because learning is determined with constructivist perspective focusing on communicative exercises facilitated by technology. Moreover, accuracy is not considered to be the only factor, as focus is shifted to language fluency. Consequently, technology plays a role of a computer learner which adapts from human learners through communicative exercises, which helps construct knowledge via communication during the learning process.

In the integrated CALL approach, the model centers upon multimedia and the internet as the means of technology. Within this approach, English teaching methods shift concern toward content-based learning, which features English for Specific Purposes (ESP) and English for Academic Purposes (EAP). This learning theory stands with a social constructivist view. As a result, language learning is moved from a cognitive to socio-cognitive perspective which focuses on the development of language through social interaction. Therefore, the principal use of technology does not only anchor communication exercises but it also does so with authentic discourse, the principal objective in which agency is seen as the focal point of language learning. Within this approach, technology plays the role of a mediating tool which helps learners achieve learning goals.

In the final and latest approach, CALL is transformed into TELL as technology can be featured by a variety of devices and tools such as mobile phones, tablets, games, and virtual worlds. This stage also witnesses the vast development among the approaches. The first and foremost transformation is from social constructivist view into connectivism, in which the central point of teaching method is about communication and interaction. This also leads to a change of the language learning perspective which sees language learning as both the socio-cognitive processes and the structural, cognitive, and adaptable processes. Interestingly, technology is normalized in this stage, and autonomy within community is the main principal objective of TELL. In the next step, we will further explore the learning theories behind TELL to gain theoretical support for our research.

**2.6.2 Learning Theories behind TELL**

According to Walker and White (2013), TELL was developed under the a connectivist framework inherited from the theories of behaviorism, constructivism, and social constructivism.

**2.6.2.1 Behaviorism**

The behaviorist view claimed that the development of language learning was through stimulus and achieved by both positive and negative reinforcement (Skinner, 1957 as cited by Chomsky, 1959). Within this framework, learning a foreign language refers to habit formation and the main objective of language learning is to use language as a native speaker. Therefore, behaviorism only centers the training of listening comprehension, accurate pronunciation, and letter recognition and reproduction. Within the behaviorist paradigm, the principle is that machine is seen to contain the knowledge and it delivers the knowledge to learners in terms of drill and practice applications or small chunks which need to be

reinforced frequently. Based on this learning theory, the machine helps learners learn by using drill and practice programs focused on the exercises like multiple-choice, short answer questions, gap-filling, re-arranging words. Learners can learn and revise their understanding independently as well as feel more secure by those mentioned activities. By all appearance, the knowledge is right there for learners, safe and sound, and what learners need to do is practice these exercises for learning to occur. The machine plays the role of an instructor to the students (Walker & White, 2013) in behaviorist paradigm.

**2.6.2.2 Constructivism**

In the constructivist framework, teaching follows a communicative language teaching (CLT) approach (Richards & Rodgers, 2014) which identifies communication and interaction as the main goals of learning language. This model focuses on learners construction of knowledge through communication and interaction. This approach is based on the theory of constructivism as suggested by Piaget (1964), that learners can best develop their learning when they can construct their knowledge by themselves from the existing knowledge which they have gained. Consequently, language learning involves the development of linguistic skills, communication, and interaction rather than simply constructing texts through drills and practice as guided by the behaviorism theory. Within constructivist paradigm, the principle is that machine no longer helps learners to learn by its ready-made knowledge but by allowing learners to teach, learn, and share so that learners can construct knowledge by themselves. Based on this learning theory, the software used for teaching is not only the software used by teachers but also by learners so that both teachers and learners can create activities for each other to construct knowledge together. Learning within this paradigm focuses on the learners ability to develop knowledge. Therefore, Walker and White (2013) simulated the role of the machine in constructivist perspective to be the learner in learning process.

**2.6.2.3 Social Constructivism**

The social constructivism originated from Vygotsky s (1978) learning theory claims that learning is shaped through social interaction before being formed by the individual aspect. With social interaction, the goal of language learning is to engage language learners in communication activities in which all factors such as text construction, linguistic skills, and interaction are combined to build up a successful language learning process. Vygotsky (1978) stated that learning is triggered by the zone of proximal development (ZPD) which refers to the connection between the ability and knowledge of learners with those of others who have more knowledge

and ability. This statement can be understood as learners learn and improve their knowledge through interacting and communicating with those who have more expertise than the learners themselves. The social constructivist approach was also reflected in Krashen s (2013) input theory in which comprehensible input plays an important role in language acquisition as learners need to acquire language through receiving the input from hearing and reading. However, this input must be at a higher level than learners knowledge and ability. Under the social constructivist framework, machines were acknowledged as tools" which facilitate communication and interaction between people. Based on the social constructivist paradigm, the machine can be used to teach, be taught, and facilitate learning process. Word-processing programs or editing-software are some examples of the tool" for learning.

**2.6.2.4 Connectivism**

The theory of connectivism suggested by Siemens (2005) explains that learning is the process which involves teachers, learners, community, and technology together in a process. Within connectivist theory, learners learn with various ideas and opinions stored in machines, and with that stored knowledge, machines play the role of tutor". Therefore, it can be implied that the behaviorist paradigm is reflected in connectivism. The constructivist paradigm is also reflected in connectivism. This statement can be supported by the assumption of Discourse Community mentioned by Walker and White (2013) relating learning to the glue which connects people in a community together. Within this perspective, an overlooked component which should be involved is the environment. The environment includes both physical and virtual entities and it is the place in which people can learn through developing the relationship with each other when they share similar interests. Since the discourse community relates the learners shared knowledge in the connected community and involves the transformation of learners knowledge into their construction of knowledge through the action of building relationship and communicating during the learning process, it can be implied that this assumption is connected with the constructivist perspective discussed by Piaget (1964). Finally, if the knowledge was stored in machines, then it can be inferred that people may use the machine to store their knowledge as well. If so, the machine can also be seen as a technological resource facilitating the learning process. Therefore, it can also play the role of a tool" as assumed in the social constructivist paradigm as mentioned above.

Walker and White (2013) utilized the Communities of Practice to explain the phenomenon in which learning is connected with practicing communities. Within this aspect, the learners who belong to a community learn by engaging in the community. This can be implied that the more learners who engage in the community, the more their expertise is developed. Learners may need to be supported by the members whose expertise is far beyond what learners currently know. However, previous researchers argued that conventional classrooms could not apply this concept for learning because were no such expert members the classrooms to help learners to go beyond what they already known. Fortunately, with the development of technology, learners can easily find an authentic community for learning and practicing with expert members in the community through the internet. They can use Facebook, Youtube, and Google to find support from the community for their learning. Therefore, it can be seen that technology plays a role of providing resources within the learning environment.

Based on the discussion mentioned above, it can be seen that this assumption is in line with the view of social-constructivism as it likely refers to the zone of proximal development by Vygotsky (1978) and the comprehensible input by Krashen (2013). Siemens (2005) suggested some learning principles guided by the connectivist paradigm, including learning determination, learning objectives, learning method, and learning process. In determining learning, connectivism suggests that learning is to understand connections in fields, ideas, and concepts. In determining learning objectives, the learner s ability to know more is very important. It s more crucial than learner s ability to know what. Therefore, the activities of learning must be accurate and the knowledge base needs to be updated continuously.

Additionally, learning should be guided by various opinions and connection which should be focused, nurtured, and maintained. Within the connectivist paradigm, learning process is conducted and associated with non-human appliances. The process should be continual and connect sources of information and specific nodes. Moreover, during the learning process, learners should be allowed to learn from reality and to decide what to learn. There are not only computers which can serve the language learning process, but other technological devices such as mobile phones, smartphones, and tablets can also play the same role, which highlighted the need of TELL in providing principles and guidelines for the learning process through technology.

Based on the above discussion, it can be concluded that TELL is constructed under the connectivist paradigm in which foundational theories and

principles of behaviorism, constructivism, and social constructivism are reflected. Therefore, technology can play many roles in the learning process. It can be the instructor to teach learners, the learner to learn from learners, or the tool to facilitate learners in the learning community with both physical and virtual entities.

## 2.7 Conceptual Framework for Developing the Lessons

The conceptual framework of this study was constructed based on the connectivist paradigm guided by TELL (Walker and White, 2013). The framework provided a clear concept of teaching English speaking skill in three stages: (1) pre-speaking, (2) while-speaking, and (3) post-speaking. Figure 2.14 s the conceptual framework for developing the lessons.

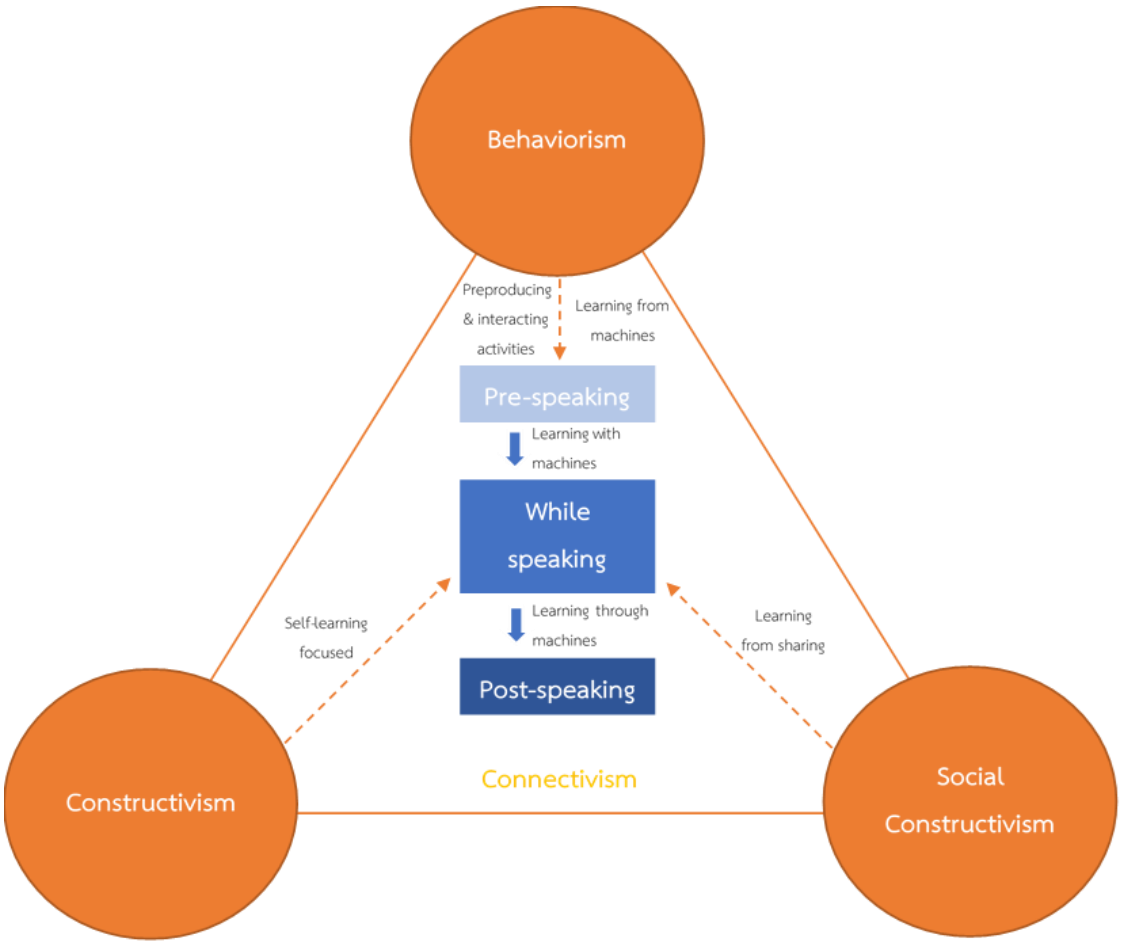



**Figure 2.14 Conceptual Framework for Developing Lessons**

### 2.7.1 Pre-Speaking

In pre-speaking stage, the behaviorism (Skinner, 1957) and connectivism (Siemens, 2005) theories should be applied in terms of warm-up and action-based

activities. Within this stage, the students learn to form habits through reproducing and interactive activities assigned in group work in which machines play the role of instructors. Accordingly, warm-up activities should be useful for students to recall and find ideas and knowledge related to the given tasks to share and support each other in their group in achieving the learning objectives. Apart from warm-up activities, the students can be taught with linguistic knowledge by a machine to have a bottom-up understanding before practicing.

**2.7.2 While-Speaking**

The while-speaking stage should be associated with constructivism (Piaget, 1964), social constructivism (Vygotsky, 1978; Krashen, 2013) and connectivism (Siemens, 2005) based theories reflected in practicing activities. Within the constructivist, social constructivist, and connectivist frameworks, the students learn and construct knowledge by themselves, learn from developing connection with people in the community, learn from those who possess expertise knowledge, and learn with non-human appliances. Therefore, the students should be provided with opportunities to learn with the machine in individual work to be autonomous in learning and be assigned with paired work to be able to learn from each other and share various ideas during the learning process. Students should also learn with the machine, and then in paired work activities to share their opinions with a partner. Sharing each other s experiences in speaking allowed each student to learn from others to improve English speaking skills.

**2.7.3 Post-Speaking**

In the post-speaking stage, the theories of behaviorism (Skinner, 1957), constructivism (Piaget, 1964), social constructivism (Vygotsky, 1978; Krashen, 2013) and connectivism (Siemens, 2005) are integrated into individual tasks to help the students develop their self-study ability so that they can learn and construct their knowledge of vocabulary, pronunciation, and grammar by themselves. This supports their developing existing knowledge through communication and connection with the community through a non-human appliance whose knowledge is more proficient than the students own knowledge.

## 2.8 Summary

In conclusion, this chapter has reviewed the information on English speaking skills and how Vietnamese undergraduate EFL students learn English speaking in traditional classrooms, where many problems were discovered. The chapter has also reviewed the approaches of TELL which are considered to be the theoretical

guideline for conducting the research. Significantly, the in-depth review of AI chatbot characteristics and structures together with information about how AI voice chatbots contribute to English language learning has been discussed. In the next chapter, this research continues with the  description of the research methodology.

# CHAPTER 3
# RESEARCH METHODOLOGY

This chapter is focused upon the research methodology for this study. This chapter is organized with nine sections including research context, research design, participants, instruments, data collection procedure, data analysis, trustworthiness and validity, and pilot study.

## 3.1 Research Context

The study was conducted at Can Tho University in Vietnam. Data collection was limited to the Faculty of Foreign Languages because this faculty is in charge of the education and training of English to all students, regardless of major. The Vietnamese undergraduate EFL students who participated in this research came any non-English degree based academic programs, with any major. These students were enrolled in the English 2 level course. Each week, the students were required to be in the classroom for nine hours to learn English basics. The Faculty was responsible for screening the students by using TOEIC Bridge as the placement test to enter the course. The main type of training is regular teaching, and the English classes occur onsite, which requires the students to be in the classroom to study. The English 2 course is taught in traditional classrooms in which teacher gives lectures while learners listen and take notes. More recently, learners have been also allowed to use smartphones to facilitate their learning process. This was an advantage for this study because we could employ Andy English Bot which supports Android and iOS to integrate with in-class teaching. Therefore, with available smartphones, the target students could easily access the chatbot to speak English.

## 3.2 Research Design

This study applied an experimental design with one independent variable and two dependent variables. The independent variable was the use of an AI voice chatbot in teaching English speaking skill. The dependent variables included (1) the students speaking test scores from the pre-speaking and post-speaking tests and (2) their opinions about using an AI voice chatbot. A mixed method design was conducted because this study contained two factors: (1) the between-subjects factor

which was the students speaking in the traditional classrooms and those speaking with an AI voice chatbot; and (2) the within-subjects factor which was the measurement before and after the intervention. According to Bordens and Abbott (2011), a mixed method design is one that includes a between-subjects factor and a within-subjects factor, used to assess the effects of the variables which cannot be controlled within subjects so as to maintain statistical advantage for studying the remaining variables. For quantitative design, the researcher employed the pre-speaking and post-speaking tests and a survey questionnaire to collect numerical data of the students participating in the experiment. With the pre-speaking and post-speaking tests, the researcher would like to compare whether the interventions led to significant differences in learning. The survey questionnaire would help to explore the students opinions of the intervention. For the qualitative design, the researcher used a list of interview questions to collect descriptive data from the participants. Based on the participants responses to the interview questions, the researcher provided an in-depth investigation on the students opinions about the intervention. The research was designed within six steps starting from instrumental design, then the pilot study, the pre-speaking test, the in-class teaching, the data collection, and finally the data analysis. In the first step, the researcher designed research instruments and instructional instruments to facilitate the experiment. After that, a pilot study was conducted to assess the initial plan of the experiment and evaluate the accuracy and validity of the instruments. In the third step, the researcher employed the pre-speaking tests to evaluate the students English speaking ability before the intervention. The in-class teaching was conducted in the fourth step so that the researcher could have an intervention in the learning process of the students. In the fifth step, the data from the speaking tests together with the students responses to the questionnaire and interview were collected for data analysis.

## 3.3 Participants

In selection of the participants, the purposive sampling technique introduced by Schreiber and Asner-Self (2011) was employed. The participants were assigned to two different groups which were the control group and the experimental group. Sample size was an initial step to discover the sufficient number of the necessary participants. Schreiber and Asner-Self (2011) stated that selecting sample size can be conducted by analyzing certain statistical measurements or by following the an agreed upon rule set. Accordingly, the current study consisted of 60 students who

were the non-English majored undergraduate students and were studying English 2 in Can Tho University in Vietnam. As the courses of English 2 are open every trimester in Can Tho University, it was convenient to select the target students who enrolled in this course. Moreover, the homogeneity of the students was ensured because it depended on their English placement test score which decided by testing to the L2 level of fluency. Regarding the placement test score, the Rector of Can Tho University issued the Decision No. 3003/QD-DHCT on July 22nd, 2021 regulating the course exemption and certification of course completion in academic programs for regular undergraduate education and training. Within this Decision, the TOEIC bridge was taken as a placement test to consider if a non-English majored undergraduate student was permitted to be exempted from a General English course including English 1, 2, and 3 or he/she was required to take some of the General English courses. Within this regulation, if the students TOEIC Bridge score is from 90 to 100, they will be exempted from taking the General English courses including English 1, 2, and 3. If their score is under 89, then the Faculty of Foreign Languages will decide if the students will be placed in either English 1, 2, or 3 based on their overall TOEIC Bridge score. In consideration of the overall scores, English 1 is required for those who scored between 30 and 42. English 2 is mandatory for the students who scored from 43 to 75. English 3 is mandatory for those who scored from 76 to 89. Table 3.1 illustrates the standard scale for evaluating English proficiency. Accordingly, the students English level was equivalent to A2 when they participated in the research.

**Table 3.1. The Standard Scale for Evaluating English Proficiency**

| Standard scale for 6-level of English proficiency used in Vietnam | Equivalent with CEFR standard | English levels A, B, C | IELTS | TOEIC (L&R) | TOEIC Bridge | TOEFL IPT | TOEFL CBT | TOEFL iBT | Cambridge Tests |
|---|---|---|---|---|---|---|---|---|---|
| Level 1 | A1 | A | | 120-220 | 30-42 | | 60 | | 100-120 KET |
| Level 2 | A2 | B | 4.0 | 225-445 | 43-75 | 360-449 | 96 | 30 | 120-140 KET<br>120-140 PET |
| Level 3 | B1 | C | 4.5-5.0 | 450-595 | 76-89 | 450-499 | 133 | 31-45 | 140-150 KET<br>140-160 PET<br>140-160 FCE |
| Level 4 | B2 | | 5.5-6.5 | 600-845 | 90-100 | 500-589 | 173 | 46-93 | 160-170 PET<br>160-180 FCE<br>160-180 CAE |
| Level 5 | C1 | | 7.0-7.5 | 850-940 | | 590-649 | 213 | 94-109 | 180-190 FCE<br>180-200 CAE<br>180-200 CPE |
| Level 6 | C2 | | 8.0-9.0 | 945-900 | | 650-677 | 250 | 110-120 | 200-210 CAE<br>200-230 CPE |
| | | | Top Score 9 | Top Score 990 | Top Score 100 | Top Score 677 | Top Score 300 | Top Score 120 | |

*(Decision No. 3003/QD-DHCT on July 22nd, 2021 by the Rector of Can Tho University).*

## 3.4 Instruments

This study utilized a variety of instruments to conduct the experiment. The instruments were divided into two types: research instruments and instructional instruments. The research instruments included the pre-speaking and post-speaking tests, the questionnaire, and the semi-structured interview while the instructional instruments included an AI voice chatbot, a learning website, a speaking rubric for assessment, and finally, lesson plans for the control group and the experimental group.

### 3.4.1 Research Instruments

Pre-speaking and post-speaking tests, the questionnaire and the semi-structured interview were the primary research instruments. Accordingly, the pre-speaking and post-speaking tests were used for evaluating the participants' English speaking skill before and after the experiment. The questionnaire and the semi-structured interview were utilized to explore the participants opinions about using an AI voice chatbot for improving English speaking skill.

#### 3.4.1.1 The Pre-speaking and post-speaking Tests

The pre-speaking and post-speaking tests (See Appendix A) were developed based on the CEFR s principles for assessment to evaluate the students speaking skill. Those principles suggest that the assessment should be based upon five aspects, including context, purpose, linkability, production, and standards. For the context, the speaking tests were designed for the non-English majored undergraduate EFL students who learn English 2. The purpose of the tests was to evaluate the students' speaking skill based on two criteria, fluency and accuracy. The test results were linkable because they might help examiners to predict the students' English speaking skill. Additionally, the test questions and topics could be used and re-used for many times to serve for English speaking assessment, which met the principle of production. Finally, the tests structure was standardized as it was designed based on the IELTS speaking test structure developed by the British Council, consisting of three parts. Part 1 included ten questions in which every two questions referred to one familiar topic such as food, study, hometown, daily activities, or family and friends. Part 2 included five topics for individual discussion in which the students talked about one selected topic among the five such as books, people, country, exercise, or festival. Part 3 featured two-way discussion and contained ten questions. Every two questions referred to one topic in part 2.

The IELTS speaking test structure was adopted for this research because it was investigated and confirmed as practical for scoring and interpreting

the results (Karim & Haq, 2014). Besides, the reliability and validity of the IELTS speaking test were investigated and confirmed by Li (2019) who reported that the test could assess candidates speaking ability in terms of fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation.

In addition, due to the observation that test takers scores may not necessarily demonstrate their true competence, multiple other measures may help to provide more reliable and valid assessment than a single measure. As a result, three speaking parts were utilized to assure test reliability and validity. It is worth mentioning that in speaking part 1, the students were asked to answer only one topic, and in part 3, only two questions related to the topic the students selected in part 2 were asked.

The speaking test pack containing 25 questions was provided to three experts for validity evaluation and all questions were evaluated with the score of +1 by the experts (see Appendix N). This suggested that the questions clearly measured the students speaking skills in terms of accuracy and fluency.

**3.4.1.2 The Questionnaire**

A questionnaire was utilized in this study as it allowed anonymous responses, to improve honesty of the feedback. The questionnaire was developed based on Bordens and Abbott s (2011) framework (See Appendix B). Within this framework, three features were considered before designing the questionnaire. First, keep the survey short to not overburden subjects. Second, provide a broad enough range of questions to keep track of the focused phenomenon. Third, responses are focused to avoid extraneous information. Accordingly, the questionnaire designing process started with defining the topic of this study to have clear and concise definitions to avoid ambiguity when interpreting the data. After that the questionnaire was developed, consisting of three parts including demographic information, closed-ended items and open-ended items. There were three reasons for this. First, the demographic information should be used to investigate the participants background before the intervention. Second, if only closed-ended items were provided in the questionnaire, misunderstanding might occur because different items might be understood differently between subjects. Finally, with open-ended questions, the participants could communicate more accurate understanding. However, open-ended questions needed to be limited in number to prevent fatigue of the subjects.

The demographic section included six questions which related to the participants general information: gender, age, major, year of study, English

entrance score, and technological experience to explore the participants demographic information. The demographic information set was used as a predictor variable.

The second part contained five close-ended questions to explore the participants opinions on their improvement in English speaking skill after the experiment. The close-ended items were employed for ease in summarizing and analyzing responses. The items in the second part were classified into two aspects – fluency and accuracy. Accordingly, questions 7-8 were to explore the participants' opinions on improving English speaking skill in terms of fluency, reflected by the factors such as hesitations or pauses, and their frequency of using interjections such as ah..., uhm..., oh.... Questions 9-11 explored the aspect of accuracy in terms of pronunciation, grammar, and vocabulary.

The third part allowed free writing responses for the participants to share opinions or suggestions on using an AI voice chatbot to improve English speaking skill. The open-ended items were used to allow the participants to elaborate their further opinions. The questionnaire was evaluated by experts who were the lecturers in the field of English speaking and familiar with using technology in language learning to ensure the reliability and validity.

A five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) following the framework of Cohen et al. (2007) was utilized to measure the students statements on the questionnaires. The five-point Likert scale is a way to measure participants opinions in the research and it was widely used for quantitative data collection in previous studies by (Each & Suppasetseree, 2021; Mukhallafi, 2020; Thao et al., 2019). Moreover, the five-point scale is simple to understand and rapid to answer. Actually, five-point scale was reported to be user friendly (Bordens & Abbott, 2011). Notably, Bordens and Abbott (2011) also suggested the importance of organizing the questionnaire items into a coherent and easy-to-follow format. The questionnaire items were translated into Vietnamese so that the participants easily respond.

In validating the questionnaire items, the researcher delivered all eleven questionnaire items to three experts for their evaluation. As a result, all questionnaire items were rated +1 by all three experts (see Appendix N), which meant that the questionnaire item clearly measured the objectives of the research.

The students responses to Part 2 of the questionnaire referred to their opinions on using the AI voice chatbot and were interpreted with ordinal scale ranked by the rules of lesser than" of greater than" instructed by Cohen et al.

(2007). Since the questionnaire items in Part 2 were ranked in the order of (1) strongly disagree, (2) disagree, 3 (neutral), 4 (agree), and 5 (strongly agree), the responses to such items could also be interpreted with an ordinal scale in association with the rules of  lesser than" of  greater than" as long as the interpretation complied with the rule:  crude data can yield only crude interpretation" (Cohen et al., 2007). Accordingly, the responses to each questionnaire item were interpreted by adapting the interpretation framework by Banditvilai (2016) as  strongly disagree" if the mean score was lesser than 1.51 but greater than 0.99,  disagree" if lesser than 2.51 but greater than 1.50,  neutral" if lesser than 3.51 but greater than 2.50,  agree" if lesser than 4.51 but greater than 3.50 and  strongly agree" if lesser than 5.10 but greater than 4.50. Table 3.2 shows the interpretation of the questionnaire responses.

**Table 3.2. The Interpretation of Points in the Questionnaire Part 2**

| Mean scores (M) scale | 1.00-1.50 | 1.51-2.50 | 2.51-3.50 | 3.51-4.50 | 4.51-5.00 |
|---|---|---|---|---|---|
| Interpretation | Strongly disagree | Disagree | Neutral | Agree | Strongly agree |

**3.4.1.3 The Semi-Structured Interview**

The interviews were conducted to investigate more opinions from twelve participants among the thirty who were in the experimental group because an individual interview might allow the participants to share their opinions more completely. A semi-structured interview technique was applied as this was more open to respondents and they could fully share their opinions on specific issues (Cohen, Manion, & Morrison, 2007). For the semi-structured interview, the informant sampling method suggested by Schreiber and Asner-Self (2011) was utilized as the researcher would like to purposively select and talk with twelve participants, six of whom scored the highest and six of whom scored the lowest in the post-testl. The participants  highest and lowest post-speaking scores suggested that these participants appeared to be distinct after taking the intervention in the experimental group. Therefore, it was predicted that they could share significant opinions after practicing English speaking with an AI voice chatbot.

The semi-structured interview (See Appendix C) was constructed following the non-directive framework suggested by Cohen et al. (2007). The semi-structured interview contained a smaller size of the sample and more open questions. Within the non-directive framework, the respondents felt free to share their opinions based on a few set questions, and the researcher could also prompt

for clarification on items. Accordingly, the semi-interview was conducted with seven guided questions to explore the students  process and opinions about using an AI voice chatbot to improve English speaking skill so that sufficient information could be collected for interpretation and exploration. These guided questions aimed at exploring the participants  practicing process, their opinions on speaking improvement in terms of accuracy (grammar, pronunciation, vocabulary) and fluency, and exploring the positive and negative aspects of an AI voice chatbot.

Zoom was employed to conduct the interviews to improve flexibility in terms of time and place for both the interviewees and the researcher. The interviewees were informed that their voice would be recorded during the interview to serve for transcription and interpretation.

The interview questions were also checked and validated by the experts to ensure their reliability and validity. Accordingly, the interview questions including seven guided questions were sent to three experts for evaluation, and all seven questions were rated +1 by the three experts, which indicated that those seven questions could clearly measure the research s objectives (see Appendix N).

**3.4.2 Instructional Instruments**

The instructional instruments used in this study included an AI voice chatbot named Andy English Bot, a learning website, a speaking assessment rubric, and the lesson plans for the control group and the experimental group. Andy English Bot served for the students  practice of English speaking, the learning website facilitated the students to learn English speaking skill on their own, the speaking assessment rubric provided grading criteria for speaking assessment after the pretest and posttest, and the lesson plans served for instructing the students in the control group and those in the experimental group.

**3.4.2.1 The AI Voice Chatbot**

This study used an AI voice chatbot named Andy - English Speaking Bot (Andy English Bot) as a tool for the students to practice English speaking. There were several purposes for selecting Andy English Bot as the practicing tool in the experiment.

First, Andy English Bot could be found and downloaded easily on Google Play or App Store. The students only needed to visit Google Play or App Store and type the keyword  Andy". The app would appear immediately on the first search result on Google Play or App store. Thus, most of the participants who joined the research could have Andy English Bot installed on their smartphone with ease.

Second, Andy English Bot was powered by the ZTO Labs which is a trusted technological company. With the company s policies on privacy and security, it was safe for users to install Andy English Bot on their smartphone as this AI voice chatbot did not collect or share any user data with third parties and it was also free for use unless the users preferred premium features such as complicated speaking topics, more exercises allotted per day, more grammar lessons, and unlimited vocabulary learning. As this study focused on less complex oral tasks, the basic and free features provided by Andy English Bot were considered adequate for the students to use.

Lastly, Andy English Bot received many positive reviews from previous research, one of which was that this AI voice chatbot could act as a virtual English native speaker and an English tutor (Kim et al., 2019) which could facilitate users in learning English speaking, vocabulary, and grammar. Therefore, this study expected that Andy English Bot could provide authentic inputs for the target students when they practiced English speaking.

In addition, Andy English Bot could communicate with users by both text and voice chats, which was appropriate in this study because its focus was the intervention of an AI voice chatbot into teaching English speaking skill.

One drawback of Andy English Bot reported in a previous study was that it failed to remember or saved users conversations (Kim et al., 2019). However, the mentioned drawback might be seen as an advantage because the students would not get bored with repeated answers made up as a result of the chatbot s remembering previous conversations.

**3.4.2.2 The Learning Website for The Experimental Group**

A learning website was created to facilitate the students in the experimental group to study English speaking. The lessons in the website were developed based on the lesson plans for the experimental group including five lessons featured with five topics: Journeys, Interests, Entertainment, Learning, and Tourism. Each lesson had specific instructions for the students to learn by themselves. Accordingly, the students could use their smartphone to access the website with the link https://aispeaks.mobirisesite.com/ to learn these English speaking lessons during class time. The teacher could stay in the class to facilitate the students in technical training and solving technical issues. Moreover, this website also included the links for the students to download Andy English Bot and supportive apps such as Voice Recorder and Voice Access to facilitate the students learning process. During the instrumental evaluation process, the website was sent to

three experts in English Language Teaching field for their evaluating together with the lesson plans to make sure it was validated to apply in teaching.

**3.4.2.3 The Speaking Rubric**

The analytic speaking rubric (See Appendix E) was designed as an analytic rubric whose columns stood for levels of achievement and rows for criteria of assessment. The analytic rubric was used because it allowed examiners to assess the participants level of achievement based on multiple criteria. This type of rubric also allowed the designer to assign various values to different criteria to determine an overall achievement by totaling the subscores. The designed rubric was adapted from the Tohoku Fukushi University (TFU) Foreign Language Assessment Rubrics designed by TFU Language Educators Group (TFU-LEG). It was used in an examination context for the one-on-one interview test.

The rubric was adapted because it mainly served for grading learners English speaking skill in terms of accuracy and fluency. This was in line with our objectives which focused on evaluating the students English speaking skill based on their accuracy and fluency. However, some changes were made so that the speaking rubric could better facilitate the speaking assessment process. For the first update, examples for each criterium were added so that examiners would have a clearer description for assessment. For the second update, the test takers English speaking skill was classified into the three levels recommended by the Common European Framework of Reference for Languages (CEFR) standards which included: insufficient user, basic user (A1), and independent user (A2).

Actually, the TFU s speaking assessment rubric classified the test takers proficiency into five scoring levels which are 0-not able to perform", 1-inadequate", 2-needs improvement", 3-meets expectation", and 4-exceeds expectations". However, when compared their descriptions with the CEFR standards (Cambridge, 2011), this study found that those five levels appear to be equivalent to A1 and A2 levels. According to Cambridge (2011), level A1 was described in terms of learners understanding and using language satisfied the level of using everyday expressions, introducing, giving and asking about personal details, speaking slowly and clearly. The TFU-LEG s speaking rubric also used similar measures for the speaking scores achieved by test takers in terms of their ability to understand and use basic, simple, and appropriate language and expressions. To simplify the scoring process, a range from 0 to 5 points was used for evaluating the test takers levels of achievement based on two criteria, fluency and accuracy.

The speaking rubric was delivered to the experts for validation and the evaluation; results revealed that most of the criteria were evaluated to clearly help the examiners to evaluate the participants English speaking skill because the average scores of those criteria were 1.00 collected from the three experts (see Appendix N). However, criterium number 1, 13, and 20 got the average score of 0.67, which meant that two experts agreed on the possibility of those criteria in facilitating the examiners in the scoring process for the speaking tests but the remaining expert was unconvinced. Based on the pre-determined convention mentioned previously in section 3.3.1, it could be concluded that the speaking assessment rubric could help the examiners with the scoring process because no criteria was below a 0.67. However, one expert suggested simplification of the assessment rubric so that the prospective examiners would find it easy to respond when giving score to the participants speaking skill. As a result, the researchers of this study decided to make the speaking assessment rubric simpler by reducing the criteria to two main criteria which were speaking fluency and speaking accuracy. Based on the literature review mentioned in Chapter 2, it was reasonable that speaking fluency could be observed based on the speaker s pauses and hesitations while speaking accuracy could be observed based on the speaker s accurate use of vocabulary, grammar, and pronunciation. Each criteria also had five levels but the levels were not mentioned in the assessment rubric. Instead, those five levels were replaced with the band score 1-5 so that the examiners would find it simple to refer to when they gave scores for the participants speaking skill. Within this aspect, the score of 1 means not able to perform", the score of 2 means inadequate", the score of 3 means needs improvement", the score of 4 means meets expectation", and finally the score of 5 means exceeds expectations".

**3.4.2.4 The Lesson Plans**

In instructing the English speaking lessons, there was a lesson plan for teaching the students in the control group and a different one for teaching the students in the experimental group. The former lesson plan was used by a teacher in a traditional classroom of English speaking while the latter one was used to instruct the students to learn English speaking by themselves with teacher s facilitation in technical issues..

**3.4.2.4.1 The Lesson Plan for the Control Group**

The lesson plan for the control group included five units featured with five topics: Journeys, Interests, Entertainment, Learning, and Tourism taught in a traditional class. Each unit took three periods in which 45 minutes were

be spent for each. In each period, there were specific activities of teaching and learning. Normally, the teacher introduced the lessons and let students learn vocabulary and grammar from the textbook. Then some paired work or group work activities were given to students to do. After that, the teacher asked students to listen to sample dialogues from the textbook to practice listening. Next, the teacher asked students to practice English speaking with their partners while the teacher observed students speaking performance. During practice time, the teacher instructed students to take turns speaking with each other based on the textbook s suggestion to practice English speaking. Finally, the teacher gave feedback to students English speaking performance at the end of each lesson so that students could become aware of their strengths and weaknesses in English speaking for future improvement.

**3.4.2.4.2 The Lesson Plans for The Experimental Group**

In the experimental group, the lesson plan was designed with the purpose of getting the students practice English speaking with Andy English Bot to improve their English speaking skill in terms of fluency and accuracy as Kim (2016) reported that an AI voice chatbot could enhance the students English speaking skills. Accordingly, the structure and contents of the lesson plans were designed based on the CEFR s Can-Do Statement suggested by Cambridge (2011) with the application of Task-Based Language Learning method.

Regarding the Can-Do Statements, it is explained as the description of what language learners can do at different stages of their learning across five language skills: spoken interaction, spoken production, listening, reading, and writing. The overall focus of Can-Do Statements is on the learners production and their ability to engage in conversations and discussion (Cambridge, 2011). For example, an A2 learner at pre-intermediate level of English proficiency can apply simple techniques to begin, continue, and leave the conversation while a B2 learner at intermediate English proficiency level can appropriately use complex phrases or sentences to develop their conversation.

There are six target levels concerned in the CEFR: levels A1, A2, B1, B2, C1, and C2. Level A1 evaluates the learners in terms of their ability to understand and use basic expressions and phrases when communicating while level A2 additionally evaluates their ability to make sentences to communicate. Level B1 evaluates the learners in their ability to understand key points of standard input, dealing with common situations, producing simple connected texts, and using language to describe and share familiar topics. Level B2 shifts the evaluation to the

learners ability to understand complex text, fluent interaction, and production of detailed text for discussion. Level C1 and C2 drive the evaluation on learners speaking proficiency in terms of understanding contextual meaning and well-organized discussion using language.

As the subjects in this study were the non-English majored undergraduate students, the lesson plan focused on getting the students to achieve level A1 and A2 only, which meant that the students were expected to be able to understand basic topics and use appropriate phrases or sentences to talk about familiar topics..

In addition, the lesson plan was designed with five lessons but those lessons were taught from a learning website. Each lesson was separated into three periods which took 45 minutes each. There were three stages of teaching and learning which were before speaking, while speaking, and after speaking. In the first stage, students were given a task to do on the website based on the current topics of the lesson to learn vocabulary, pronunciation, and grammar. In the second stage, students practiced English speaking with Andy English Bot and focused on the correct use of vocabulary, grammar, and pronunciation when speaking. In the final stage, students took a speaking exercise with Andy English Bot and then self-evaluated their English speaking skill by looking at the chat screenshots and listening to the voice chat recordings to see how they improved throughout the learning stages before getting feedback from the teacher.

The lesson plan focused on five main topics which were journeys, interests, entertainment, learning, and tourism. The reason for choosing these topics was that they might be familiar to most of undergraduate EFL students at A2 level. These topics were chose based on the finding of Ellis (2009) that less complex tasks might enhance oral fluency.

The lesson plans contained eight survey statements and each statement was rated by the five-point likert scales from 1 to 5, respectively equivalent to strongly disagree, disagree, neutral, agree, and strongly agree. The evaluation results revealed that there were three statements which received a 100% agreement from three experts: statement numbers 3, 7, and 8. Statements 1, 5, and 6 received a 93.33% agreement from the experts. Finally, statements 2 and 4 received an 86.67% agreement from the experts (see Appendix N). The results implied that the lesson plans met the teaching objectives.

## 3.5 Data Collection Procedure

Overall, this study collected both quantitative and qualitative data to achieve the research objectives and answer the research questions. The instruments used thorough the study were evaluated by experienced experts to ensure the validity and reliability of each instrument before being applied into the experiment.

For general process, sixty participants participated in the eight-week experimental period starting from the tenth of April, 2023 with introduction sessions, orientation sessions, and pre-speaking tests after the researcher had got an approval from the Ethnics Committee for conducting the experiment. The number of participants and the term of experiment were determined based on the research design by Han (2020); Kim (2017); and Podesva and Sharma (2013). In addition, an eight-week experimental period was taken for the research to attempt to avoid Hawthorne effect which could possibly cause participants' over achievement due to their knowledge of being in an experiment (Kim, 2018).

Before participating in the research, the subjects were given an information sheet and a consent form for their review and signature to ensure that they agreed to participate in the experiment by their own free will. This process was to satisfy the requirement for ethics in human research.

The study collected the quantitative data from the students pre-speaking and post-speaking test results to analyze and find answers to research question 1. The students responses to the questionnaire were collected and analyzed quantitatively while the results from the interview were collected and analyzed qualitatively to find answers to research question 2.

### 3.5.1 Quantitative Data Collection

In collecting the quantitative data, 60 participants took a pre-speaking test at the beginning so that the researcher could have baseline data on the students English speaking proficiency before the intervention. They were then randomly assigned in two different groups: a control group and an experimental group. Each group contained 30 students.

The students in the control group learned with a textbook and spoke English in a traditional English speaking classroom while the experimental group learned on the website and practiced English speaking with Andy English Bot. Kim et al. (2019) reported that this AI voice chatbot could serve as communicative and learning tools for the students to learn English speaking.

The researcher was in charge of teaching English speaking to the students from the lesson plans in both groups. This ensured the research framework was

strictly followed. It eliminated extraneous variables which could have occured if other teachers provided instruction.

Instructional sessions on how to use Andy English Bot had been conducted before the students in the experimental group began the intervention.

During class time, all students learnt English speaking about some familiar topics such as journeys, interests, entertainment, learning, tourism, with a focus on speaking with accuracy and fluency.

The participants in the control group learned and practiced English speaking face-to-face with their partners as usual in traditional classrooms while those in the experimental group were asked to practice speaking with Andy English Bot for around 30 minutes in class for each unit following Cambridge s (2017) suggestion that about 25% of the classroom time is ideal for communicative speaking tasks. While speaking, the students were encouraged to record their voice and take screenshots of the chat sessions to provide tracking data.

After five weeks of learning English speaking, all students took a post-speaking test. For conducting the speaking tests, the researcher provided the questions to the students and recorded the responses of the students in the experimental group while the students in the control group took the speaking test in a pair following the traditional manner in which one student asked and another one answered.

The topics for the pre-speaking and post-speaking tests were similar to the topics the students learned during class time. The study used familiar topics for the students to practice English speaking so that they would not find it too difficult to speak. This decision was supported by Ellis s (2009) research finding which revealed that with less complex tasks, the students could achieve greater oral fluency and accuracy. Finally, the participants scores were graded by the three examiners from the pretest and posttest were collected for analysis.

The process of grading was conducted by three examiners, one English native speaker and two non-native speakers of English. The examiners were selected under the criteria for English speaking examiners suggested by the researcher in Appendix F. This study had adapted and adjusted the criteria from the IELTS to test examiners knowledge and skill requirements. Since this research used the English speaking tests for the students at A1 and A2 levels, the criteria for the speaking examiners in this research was more relaxed than the IELTS s criteria for speaking examiners. The speaking tests were structured following the actual IELTS speaking test structure. It is also worth mentioning that the English speaking examiners were

trained by the researcher before grading the students in the pre-speaking and post-speaking tests.

The scores of the students in the pre-speaking and post-speaking tests were graded based on the speaking assessment rubric. For grading procedure, in each speaking element, the students were scored based on two criteria – fluency and accuracy. After that, the average scores for each criterium graded by the three examiners were averaged to ensure reliability for each criterium. The total score was then summed from the average scores for each criterium to ensure consistency. The scores of the examiners were often similar from one examiner to another. No reconsideration of individual scores was requested.

The test for the students in the experimental group was online and took around 5-10 minutes while the test for the students in the control group was conducted in class traditionally. The double-blind rating method was applied to avoid raters bias when they graded the students speaking skill in both pre-speaking and post-speaking tests. The term double-blind rating" was implemented in this study based on the definition of double-blind peer review" defined by Casserly (2017) in which the status of both the author and the reviewer were kept unknown. Double blind style assessment ensures no bias from the evaluators since they do not personally know who is being evaluated.

There were some concerns on whether the process of reviewing was actually blinding based on this approach or whether the reviewers could potentially guess about the subject's work and style. This concern was reduced by preventing the reviewers from referencing other materials during the tests.

The double-blind model was applied in this study in two ways. First, the pre-speaking and post-speaking test recordings from the students were kept secret until the end of the experimental period and then mixed up together before they were delivered to the raters. This ensured there was no information available about the subjects for the pretests and posttests provided. Second, the information whether the recordings came from the control group or the experimental group was not known by the examiners. This prevented experimenter bias.

After taking the post-speaking tests, the students in the experimental group further responded to the questionnaire so that the research could collect quantitative data for analysis. Within this process, quantitative data were collected from the scores of the students in the pre-speaking and post-speaking tests and from their responses to the questionnaire after the post-speaking test.

### 3.5.2 Qualitative Data Collection

In collecting the qualitative data, a semi-structured interview was conducted under the informant sampling framework in which the researcher selected six students in the experimental group who got the highest score and six who got the lowest scores in the post-speaking test to take the interview.

Some guided questions were asked to explore in-depth information of the students opinions on using an AI voice chatbot to improve English speaking skill. Throughout this process, qualitative data was collected from the participants responses in the interview.

In summary, this study collected the quantitative data from the participants pre-speaking and post-speaking tests to answer research question 1. The quantitative data from the students responses to the questionnaire and the qualitative data from their responses to the semi-structured interview were collected and analyzed to answer research question 2. Table 3.3 summarises the research instruments and data analysis methods used in answering two research questions.

**Table 3.3. Summary of Research Instruments and Data Analysis**

| Research Question (RQ) | Instrument | Data Analysis |
|---|---|---|
| RQ1: What are the effects of an AI voice chatbot on English speaking skill of Vietnamese undergraduate EFL students? | ▪ Pretest and posttest speaking (quantitative data) | ▪ Paired sample t-test |
| RQ2: What are the opinions of Vietnamese undergraduate EFL students on using an AI voice chatbot? | ▪ Questionnaire (quantitative data)<br>▪ Semi-structured interview (qualitative data) | ▪ Descriptive analysis<br>▪ Thematic analysis |

## 3.6 Data Analysis

The data analysis process was conducted with the quantitative data and the qualitative data. Accordingly, the quantitative data included the students scores from the pre-speaking and post-speaking tests. The students scores were analyzed with the paired sample t-test provided by SPSS to answers for research question 1 What are the effects of an AI voice chatbot on English speaking skill of Vietnamese undergraduate EFL students?". After that, the data collected from the students responses to the questionnaire were quantitatively analysed with frequencies and descriptive statistic methods available in SPSS. Finally, the students responses to the semi-structured interview were qualitatively analyzed with a thematic method to provide answers for research question 2 What are the opinions of Vietnamese undergraduate EFL students on using an AI voice chatbot?".

**3.6.1 Quantitative Data Analysis**

In analyzing the quantitative data, the paired sample t-tests following Podesva & Sharma (2013) s guideline was utilized to analyze the students pretest and posttest scores. The paired sample t-tests were used in this study because there was only one independent variable which was the AI voice chatbot and two samples to be compared which were the control sample and the intervention sample. With the paired sample t-test, the *p* values derived from the comparisons were assessed in finding statistically significant differences between the two samples.

The analysis was conducting by using SPSS version 20. Furthermore, to compare the control group and the experimental group from various values, five measures were implemented. The first measure was the *p* value from comparing the mean scores of students in both control and experimental groups in the pre-speaking test to see if the students in different groups were significantly different from each other. The second measure was the *p* value from comparing the mean scores of the control group in the post-speaking test with those in the pre-speaking test to see if the students got significantly improved after the experiment. The third measure was the *p* value from comparing the mean scores of the experimental group in the post-speaking test with those in the pre-speaking test to see if the students got significantly improved after the experiment. The fourth measure was the p value from comparing the mean scores of two groups in the post-speaking test to see if the intervention was significantly superior from the control group after the experimental period. The last measure was the *p* value from comparing the mean scores growth of the control group from pre- to post-speaking test and those of the experimental group from the same process to see if the experimental period had significant effects on both control and experimental groups.

**3.6.2 Qualitative Data Analysis**

In analyzing the qualitative data, the students responses to the semi-structured interview were analyzed deductively starting from interpreting the general opinions of the students and then extracting their specific opinions on each theme. The thematic analysis method discussed by Heigham and Croker (2009) was utilized. Within the thematic analysis method, five themes were suggested by the study itself in consideration of the relevance of each theme to the research objectives. Accordingly, the themes included (1) the students practice process, (2) speaking accuracy improvement (grammar, pronunciation, vocabulary), (3) speaking fluency improvement, (4) positive aspects of an AI voice chatbot, and (5) negative aspects of an AI voice chatbot.

Within the analysis process, the first theme contained the information of the students process of practicing English speaking with Andy English Bot, the second and third themes referred to the students points of view on their improvement in English speaking fluency and accuracy (grammar, pronunciation, vocabulary) after the intervention, and the final theme focused on the positive and negative aspects of the AI voice chatbot. The thematic analysis method was reported to be appropriate for grouping data within specific themes and this method would allow for for data interpretation.

## 3.7 Trustworthiness and Validity

### 3.7.1 Trustworthiness

This current study employed the indexes of items-objective congruence (IOC) model introduced by Turner and Carlson (2003) to evaluate item validity of the instruments, including the speaking test, the questionnaire, the semi-structured interview, the lesson plans, and the speaking assessment rubric. The index of item-objective congruence (IOC) was defined by Turner and Carlson (2003) as *a procedure used in test development for evaluating content validity at the item development stage".* Within this definition, the evaluation of an independent expert panel to rate the suitability of the items used for conducting measurement of one or more objectives can be checked to obtain evidence of item validity. The IOC was conducted experts evaluating each item by ranking the item with a rating point with the values of -1, 0, or 1. The rating of 1 meant clearly measuring" while -1 meant clearly not measuring", and 0 meant unclear in measurement."

Along with the rating point system, the independent status of experts was maintained when conducting the IOC. Turner and Carlson (2003) suggested that the experts should not been informed of the target constructs which the individual items were intended to measure.

In this process, a set of items was created to measure the mentioned constructs. After the creation, each item was listed in the rows of a table and the possible measuring objectives were put in the columns. Next, the created list was distributed to the experts for rating each item by utilizing the rating point system as described above. Finally, the ratings were combined to give IOC measurement results for each item on each objective. The combined score of each item ranged from -1 to 1. Tthe score of 1 meant all experts agreed on the item validity and concluded that the item was not measuring any irrelevant items, while -1 meant the item validity was not valid as it was not measuring the hypothesized objective.

The IOC was also utilised in Each & Suppasetseree s (2021) s research for evaluating item trustworthiness and validity of their instruments. They had the forms evaluated by three experts in the field of English language teaching. Those experts were reported to have at least five-years of experience in teaching at the tertiary level. The results from the forms were confirmed to be reliable. This study also used the IOC model with the same number of evaluating experts who had equivalent qualification in the field of English language teaching to evaluate the instrument for trustworthiness.

**3.7.2 Validity**

For the validity of the instruments, a team of three experts who had more than five years of experience in the fields of technology-enhanced language learning (TELL) and English speaking took part in the evaluation process. The researcher provided the experts with the IOC evaluation forms for four instruments: (1) the questionnaire (Appendix G1), (2) the semi-structured interview (Appendix G2), (3) the speaking test pack (Appendix G3), and (4) the speaking assessment rubric (Appendix G4) so that they could provide a score of -1 (clearly not measuring), 0 (unclear), or 1 (clearly measuring) for each item based on their opinions about their validity.

The experts were trained on how to perform inter-rating assessments based on the index of item-objective congruence value proposed by Turner & Carlson (2003). Specifically, they were trained to do the ratings for each item which was then calculated into an average, which might result in -1, -0.67, -0.33, 0, 0.33, 0.67, or 1. In the case in which the average of the items was less than 0, the item would be eliminated. If the average score was between 0 and 0.33, that item was revised to make it clearer in meaning. The average score between 0.67 and 1 provided the best result because all experts agreed on the effectiveness of the item.

For the lesson plans validity, the researcher sent a survey on training needs to the three experts for evaluation (Appendix H). The survey was designed based on Laksana s Training Need Survey for Teachers in which the five-point rating scale was provided to the experts so that they could give their opinions about each item response by rating 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), and 5 (strongly agree), regarding the lesson plan meeting teaching objectives.

## 3.8 Pilot Study

A pilot study was conducted so that this research could evaluate the reliability of the instruments used in our research and predict possible issues of concern for

the experimental design. The pilot study generated data to assess general process, specific process, and problems discovered.

**3.8.1 General Process**

For the general process, the pilot study was conducted during four weeks with 30 voluntary Vietnamese undergraduate EFL students. Two units of the eight were taught during the pilot study via the website.

The students took a pre-speaking test and were trained on how to have conversations with an AI voice chatbot before they learned the English speaking lessons. During class time, the students practiced speaking with Andy English Bot and Google Diaglfow Bots on the website.

After finishing two units, the students took the post-speaking test to assess improvement in English speaking. This allowed the researcher to investigate the reliability of the test pack and the lesson plans.

After the post-speaking test, the students further answered the questionnaire and took part in an interview so that the reliability of the questionnaire and the semi-structured interview could be evaluated.

Based on the results of the pilot study, the researcher made some adjustments in the instruments to improve study design. The questionnaire which was evaluated by the experts as valid, was found be students to be easy to complete in all three sections. The semi-structured interview were assessed by the students as reasonable, and they were able to give direct answers and elaborate appropriately when asked.

**3.8.2 Specific Process**

Regarding the specific process, the pilot study was conducted within three various stages. At first, all instruments were submitted to three experts to review and evaluate before beginning the pilot study. Three experts selected had at least five years of teaching English speaking and held at least an MA degree in English studies and English education. The detailed results of the experts evaluation were reported in Appendix I.

For the second stage, there were 29 students who studied English 2 in a university participating in the experiment during the pilot study time. However, only 24 out of 29 participated the pilot study until the end. The other six students dropped out due to their overall course load. Those students participated in a pre-speaking test before they attended the four-week pilot. After that, they took a post-speaking test, answered the questionnaire, and participated in an interview with the researcher.

For the final stage, the pilot study invited three examiners to grade the participants English speaking skills to evaluate the data process. One examiner was a native speaker of English who came from America and other two examiners were non-native lecturers of English who were teaching English at a university in Vietnam. All examiners had at least five years of teaching experience and held at least an MA degree. The researchers created a Google Drive folder, uploaded all recorded audio files of the students from the pre-speaking and post-speaking tests including the grading sheets and the speaking assessment rubric in to the folder. A link was shared with the three examiners for their assessment. All audio files were put into specific subfolders labelled PARxx" in which xx" stood for the number of the subject Each subfolder contained two audio files recorded from the pre-speaking and post-speaking tests of the subjects. The examiners noted scores directly to the grading sheet without the need for downloading. After the grading process, the scores given by the three examiners were averaged. The average scores were selected as the final scores to be analyzed. The detailed results of each stage throughout the pilot study are located in Appendix J, containing three main parts: the research instrument evaluation, the pilot study, and the pilot results.

**3.8.3 Pilot Results**

The results of the pilot study are discussed below, including quantitative data and qualitative.

**3.8.3.1 Quantitative Data Analysis Results**

For the results of quantitative data analysis, the students scores from the pre-speaking and post-speaking tests were collected and analyzed using the paired sample t-test in SPSS to investigate if there were any significant differences between the Vietnamese undergraduate EFL students who learnt English speaking skill in the traditional classroom and those who used an AI voice chatbot to learn English speaking skill. The results of the speaking tests showed that there were significant differences between the pre-speaking and post-speaking tests of the students in the experimental group ($p = 0.00$).

The questionnaire responses in Q7 and Q8 revealed that more than 70% of the students believed that speaking with the AI voice chatbots helped them to improve their fluency in speaking because they could speak with less hesitation and pauses and use hedging words more appropriately.

The responses in Q9, Q10, and Q11 revealed that around 60% of the students believed in their improvement in speaking accuracy after speaking with

the AI voice chatbots as they could pronounce correctly, use appropriate sentence structures, and choose appropriate words and vocabulary.

**3.8.3.2 Qualitative Data Analysis Results**

For the results of qualitative data analysis, the interview responses revealed that most of the students believed their improvement in grammar and pronunciation was supported by the interaction with the AI voice chatbots. Some of the students reported that when they uttered a sentence, and their sentence appeared on the screen, this was useful feedback because they could see whether the grammatical structures were correct. For example, PAR03 said: *"When I spoke with the AI voice chatbot, I found that my grammar was improved very much because when I produced a sentence, that sentence appeared on the chat screen between me and the AI voice chatbot. Therefore, I knew what grammar mistakes I was making and how to overcome."*

The participants also reported that they could learn new grammatical structures after talking with the AI voice chatbots because it also used new grammar structures and new vocabulary. For instant, PAR16 reported: *"I found that my grammar had been improved because every time the AI voice chatbot talked to me, they used new grammar structures which encouraged me to learn more."*

Regarding vocabulary improvement, the participants reported that their improvement in vocabulary required both practice with the AI voice chatbots and their own knowledge of vocabulary. Actually, PAR03 mentioned: *"I think that my vocabulary has not improved much because my available vocabulary is still limited."* Even when asked if PAR09 found that his vocabulary was improved after speaking with an AI voice chatbot, he responded: *"I think no because it did not explain in Vietnamese. Therefore, I did not know the words, too."*

The students' responses in the interview also revealed that after speaking with an AI voice chatbot, they found themselves speaking with less pauses and hesitations because they could practice speaking repeatedly until the AI voice chatbots could understand them. This repetition helped them become familiar with the ideas they wished to convey which allowed them to speak more fluently. PAR11 reported: *"When I practiced with the AI voice chatbot, if I spoke with too many hesitations, then it would not understand at all. Therefore, I had to speak slowly and repeat many times so that it could understand."*

When the students were asked about their feeling about using the AI voice chatbots, they reported a range of responses. Some students reported that

they were interested in using the AI voice chatbots for improving English speaking skill because they believed that the AI voice chatbots provided a friendly interaction in English which was comfortable. The AI voice chatbots could also speak clearly and loudly enough to make the students feel like they were talking with a native speaker. PAR01 said: *"When I spoke with the AI voice chatbot, this made me feel friendly. The AI voice chatbot pronounced clearly and loudly enough, and it could also repeat many times."* The students also reported that they thought the AI voice chatbots could be their tireless friends because those chatbots could talk with them tirelessly at any time in any places. PAR05 reported: *"When I practiced with the AI voice chatbot, the thing I like best is that the restless time we could take for asking and answering."*

However, there were two things that the students did not like about using the AI voice chatbots for practicing English speaking. The first was the limitation of speaking topics, and the second was technological issues regarding the AI voice chatbots' ability to recognize Vietnamese names. Most of the students reported that they felt quite bored because they could not talk about many topics when speaking with the AI voice chatbots. For instance, PAR05 shared: *"The thing I dislike is that because the AI voice chatbot is a robot, it can only reply to me with fixed sentences. Its responses are not plentiful. I hope that there will be more topics and a variety of responses so that people can have more choices. Sometimes, the AI voice chatbot only asked for my repetition because it failed to answer me."* Other students reported that the AI voice chatbots could not recognize Vietnamese names even though they repeated it several times. For example, PAR03 said: *"The thing I like best is that I can improve my English and speak better. However, there are some sections which I speak again and again but the AI voice chatbot still could not get it."*

Based on the quantitative and qualitative results, there were three findings which could relate the research questions. First, there were significant differences between Vietnamese undergraduate EFL students who studied in the traditional classroom and those who studied using an AI voice chatbot in terms of speaking fluency and accuracy, demonstrated by the *p* values of 0.00 in the paired sample t-test when comparing the English speaking scores of the students in the pretest with the posttest. Second, the AI voice chatbots were found to help the students improve their English speaking skill because most of them reported that after the experiment period, they could speak English with less hesitation and pauses and use interjections appropriately. Moreover, they also reported that speaking with

an AI voice chatbot helped them speak more accurately in terms of pronunciation, grammar, and vocabulary. Finally, most of the students expressed positive opinions on using the AI voice chatbots while some shared constructive opinions.

**3.8.4 Problems Found**

During the process of the pilot study, there were some problems found in terms of the speaking tests, the practice time requirement, and technical issues.

The first problem related to the questions for the speaking tests in Part 1 and Part 3. In Part 1, the questions in the topics of food, study, and daily activities were unfamiliar to the students because there were some words used at a higher level than the students' current level of fluency. Therefore, these questions were adjusted appropriately. In Part 3, the structures of some questions related to the topics of people, country, exercise, and festivals appeared too complicated for the students because these questions were designed with an advanced grammatical structure. These questions were downgraded to a simple structure so that these questions would be more comfortable for students at this level.

The second problem related to the time the students were assigned to practice and record audio files of their speaking with an AI voice chatbot. At first, the students were expected to record their speaking session with an AI voice chatbot after class for about ten minutes per week. They later believed this request was too burdensome because they had duties from other coursework. To solve this problem, the reseacher decided to reduce practice time in English speaking with an AI voice chatbot to class time only. They could then voluntarily provide screenshots of speaking sessions with an AI voice chatbot if they have time after class, without it being a course requirement.

The final problem related to the Google Diagflow bots utilized in the pilot. At the beginning, we decided to use both Andy English Bot (Andy) and Google Diagflow bots (namely Peter and Elsa) to speak with the students. Therefore, the students had the choice to practice and speak English with either Andy, Peter, or Elsa. However, some students did not have laptop accessibility, which was required for two of the programs. As a result, we decided to utilize Andy as the primary AI voice chatbot for the students to practice English speaking because all of the students possessed a smartphone compatible with Andy. Moreover, Peter and Elsa were reported to have limited topics for students to practice English when compared to Andy. This was a second reason the researcher decided to use Andy in the study. The final problem was that both Andy English Bot and Google Diagflow bots failed to recognized the students Vietnamese names. To solve this problem, the students

should consider using an English nickname when speaking with Andy rather than their Vietnamese name.

## 3.9 Summary

In summary, this section has discussed the research methodology and design. The next section will delve into results of the study and discussion of the findings.

# CHAPTER 4
# RESULTS AND DISCUSSION

This section will review the results of the study, and discuss the meaning thereof to address the two research questions: (1) What are the effects of an AI voice chatbot on English speaking skill of Vietnamese undergraduate EFL students? and (2) What are the opinions of Vietnamese undergraduate EFL students on using an AI voice chatbot?

## 4.1 Results

The experiment was conducted within an eight weeks timeframe with thirty participants in the control group and thirty in the experimental group. The pre-speaking and post-speaking tests and the questionnaire items were used as the quantitative data while the semi-structured interview was employed as the qualitative data. Overall, the students scores from the pre-speaking and post-speaking tests in the experimental group showed significant difference with the *p* value of 0.00 ($p<0.05$) and all students agreed that using the AI voice chatbot could improve their English speaking skill, demonstrated by the total mean score of 3.90 ($M=3.90$) after interpretation. The focus of the results of the study will be described in terms of the following two questions:

(1) What are the effects of an AI voice chatbot on English speaking skill of Vietnamese undergraduate EFL students?

(2) What are the opinions of Vietnamese undergraduate EFL students on using an AI voice chatbot?

Both quantitative and qualitative data from the pretest and posttest, the questionnaire, and the interview will be reported. The pretest and posttest results will be compared to discover the effects of an AI voice chatbot on English speaking skills of Vietnamese undergraduate EFL students. The questionnaire and the semi-structured interview responses will be used to explore the students opinions on using the AI voice chatbot.

### 4.1.1 The Effects of an AI Voice Chatbot on English Speaking Skill of Vietnamese Undergraduate EFL Students

In investigating the effects of an AI voice chatbot on English speaking skill of Vietnamese undergraduate EFL students, the quantitative results were analyzed.

Accordingly, the scores of the students in the pretest and posttest were analyzed by using paired sample t-tests.

The detailed results are found in Table 4.1 and revealed the students improvement in English speaking skill generally and their speaking accuracy and fluency particularly. Some significant differences between the student s English speaking improvement in the control group and in the experimental group were also found in the table.

**Table 4.1. Results of the Paired sample T-tests**

| Paired Samples Test | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | | Lower | Upper | | | |
| Pair 1 | PretestCG - PretestEG | -1.0667 | 3.0954 | .5651 | -2.2225 | .0892 | -1.887 | 29 | **.069** *p>0.05* |
| Pair 2 | PosttestCG - PosttestEG | -1.5667 | 3.2129 | .5866 | -2.7664 | -.3669 | -2.671 | 29 | **.012** *p<0.05* |
| Pair 3 | PretestEG - PosttestEG | -.8167 | .4676 | .0854 | -.9913 | -.6421 | -9.565 | 29 | **.000** *p<0.05* |
| Pair 4 | PretestCG - PosttestCG | -.3167 | .5676 | .1036 | -.5286 | -.1047 | -3.056 | 29 | .005 *p<0.05* |
| Pair 5 | GrowthCG - GrowthEG | -.5000 | .6565 | .1199 | -.7452 | -.2548 | -4.171 | 29 | **.000** *p<0.05* |
| Pair 6 | AccuracyPretestEG – AccuracyPosttestEG | -.2833 | .2780 | .0508 | -.3872 | -.1795 | -5.582 | 29 | **.000** *p<0.05* |
| Pair 7 | FluencyPretestEG - FluencyPosttestEG | -.4333 | .3198 | .0584 | -.5528 | -.3139 | -7.421 | 29 | **.000** *p<0.05* |

In pair 1, comparison of the students scores in the pretest for both groups showed no significant difference between the control group and the experimental group, demonstrated by a p=0.69 (p>0.05). This measure demonstrated that students were equivalent in English speaking before the intervention.

In pair 2, the students post-test scores were compared which revealed that the scores of students in the experimental group significantly different from the control group, demonstrated by the p value of 0.012 (p<0.05).

In pair 3, students in the experimental group showed significant improvement in their English speaking scores in the posttest compared with the pre-test with the p value of 0.000 (p<0.05).

In Pair 4, those in the control group also improved their English speaking scores in the posttest with p=0.005 (p<0.05). These p values suggest that after eight

weeks of learning English speaking skills, the students in both groups could improve their English speaking skill.

However, as demonstrated by pair 5, the students who used an AI voice chatbot to practice English speaking in an AI-aided classroom were had significantly more improvement than the control group, as demonstrated by the p value of 0.000 ($p<0.05$).

In pair 6, the scores of speaking accuracy compared pre-test and post-test of students in the experimental group. The results demonstrated that speaking accuracy was significantly improved after use of an AI voice chatbot to learn English speaking, suggested by the p value of 0.000 ($p<0.05$).

In pair 7, the scores of the students in the experimental group were compared pre-test and post-test for fluency. The results suggested that students fluency was significantly improved after they used an AI voice chatbot to learn English speaking, indicated by the p value of 0.000 ($p<0.05$).

**4.1.2 The Students 'Opinions on Using an AI Voice Chatbot**

The questionnaire responses were analyzed using descriptive and frequencies statistical method. The results from the semi-structured interview were qualitatively analyzed with the thematic analysis method.

**4.1.2.1 Questionnaire Responses**

The results of the pre-test and post-test responses were analyzed using the frequencies and descriptive statistical method to provide quantitative results. The questionnaire was used to collect two sorts of information. The first was the students' demographic information and the second was the students' opinions on using an AI voice chatbot to learn English speaking skill based on the questionnaire.

***A. The Students' Demographic Information***

The demographic information of the students included their gender, age, major, year of study, placement test score, and technological experience. In line with the demographic information, the results of the students' gender, major, and technological experience will be illustrated in percentages while the results of age, year of study, and English placement test score will be presented using mean scores.

**Table 4.2. The Students' Personal Information**

| | Students' Personal Information | | | | | |
|---|---|---|---|---|---|---|
| Gender (%) | | Age (Mean) | Non-English Major (%) | Year of study (Mean) | English placement test score (Mean) | Technological experience (%) |
| Male | Female | | | | | |
| 13 (43.3%) | 17 (56.7%) | 20.67 | 100% | 2.73 | 53.8 | 100% |

In Table 4.2, there were 30 students participating in the study. Among the students, there were 13 males and 17 females which were 43.3% and 56.7% of the sample, respectively. The students' average age was around 21 and most were in the 2nd or the 3rd year of study. Their average English placement test score in the TOEIC Bridge test was of 53.8, which was at the A2 level. All students came from non-English majors and had technological experience prior to the intervention.

***B. The Students' Opinions on Using an AI Voice Chatbot***

The five items in Part 2 of the questionnaire were analyzed by the use of a descriptive statistic method to find out the students' opinions on using an AI voice chatbot to learn English speaking skills. The analyzed results were interpreted based on Table 3.2 (The Interpretation of Points in the Questionnaire Part 2) explained in Section 3.4.1.2 previously. The interpreted results are provided in Table 4.3.

**Table 4.3. Students' Opinions on Using the AI Voice Chatbot**

| Descriptive Statistics | | | | Interpretation |
|---|---|---|---|---|
| | N | Mean | Std. Deviation | |
| Q7 - After practicing with the AI voice chatbot, I think I could speak English without making too many pauses and hesitations. | 30 | 3.73 | 0.74 | Agree |
| Q8 - After practicing with the AI voice chatbot, I think I could speak English with appropriate hedging words such as uhm... ah... oh.... | 30 | 4.07 | 0.74 | Agree |
| Q9 - After practicing with the AI voice chatbot, I think I could produce correct pronunciation when I spoke English. | 30 | 3.87 | 0.90 | Agree |

**Table 4.3. Students' Opinions on Using the AI Voice Chatbot (Cont.)**

| Descriptive Statistics | | | | Interpretation |
|---|---|---|---|---|
| | N | Mean | Std. Deviation | |
| Q10 - After practicing with the AI voice chatbot, I think I could use appropriate sentence structures when I spoke English. | 30 | 3.90 | 0.71 | Agree |
| Q11 - After practicing with the AI voice chatbot, I think I could use appropriate words and vocabulary when I spoke English. | 30 | 3.93 | 0.78 | Agree |
| Valid N (listwise) | 30 | | | |
| **Total:** | | **3.90** | **0.77** | **Agree** |

In Table 4.3, the results showed that all students agreed with all five items in Part 2 of the questionnaire with the total mean score of M=3.90 and SD=0.77. The results indicated that the students agreed that they had improved English speaking skill in terms of fluency and accuracy after practicing English speaking with the AI voice chatbot. Among the five items, the highest mean score fell into item 8 - "*After practicing with the AI voice chatbot, I think I could speak English with appropriate hedging words such as uhm... ah... oh....*" which received a score of 4.07 out of 5.00. Item 7 " *–After practicing with the AI voice chatbot, I think I could speak English without making too many pauses and hesitations."* received the lowest mean score of 3.73 while the items 9, 10, and 11 received a medium mean score of 3.87, 3.90, and 3.93 respectively. However, the score of 3.73 did not affect the interpretation. All mean scores obtained 3.50, or greater, which was interpreted as "agree" based on the definitions provided in Table 3.1 mentioned in Section 3.4.1.2 in Chapter 3 of this document. The results suggested that all students agreed that using the AI voice chatbot could help them improve English speaking skill in terms of fluency and accuracy.

#### 4.1.2.2 Semi-Structured Interview Responses

The semi-structured interviews were employed to provide greater insight of the students opinions about using an AI voice chatbot to learn English speaking skills. There were 12 students selected based on their post-test scores, six of whom were in highest scoring group and six of whom were in the lowest scoring group. This selection relied on the premise that these students might have significantly different opinions on the use of an AI voice chatbot to learn English speaking skills due to their post-test results. The responses of the students in the

interview were collected and interpreted by using the thematic method. The themes were organized into (A) the students practice process, (B) the improvement in speaking fluency, (C) the improvement in speaking accuracy (grammar, pronunciation, vocabulary), (D) the positive aspects of the AI voice chatbot, and (E) the negative aspects of the AI voice chatbot.

***A. The Students' Practice Time with AI Voice Chatbot***

All 15 participants were asked how often they practiced English speaking with the AI voice chatbot every week. Their answers revealed that most spent time speaking with the AI voice chatbot during fee time, but the length of time between groups was different. Some spent 10 to 15 minutes to speak every week while others spent 30 to 60 minutes or more to practice and record their speaking sessions.

> **PAR05:** *"Normally I spoke around 3-4 times a week when I was free. It helped me to kill time and learn more. I usually spoke for around 10-15 minutes each time."*
>
> **PAR06:** *"I practiced English speaking with the AI voice chatbot everyday for around 30-60 minutes depending on each day availability."*

***B. The Improvement in Speaking Accuracy***

When the participants were asked if they thought their grammar had improved after speaking with the AI voice chatbot, most of them found that they could improve grammar more effectively after speaking with the AI voice chatbot. This was in part because they could look at the chat screen to evaluate their use of grammar within an uttered sentence. Moreover, the AI voice chatbot used correct grammar structures which could be used as a model for students to learn from.

In terms of accurate pronunciation in speaking, the students reported that most found that they could speak with better pronunciation because they could listen to the native voice of the AI voice chatbot to practice pronunciation multiple times. The visual display on the chat screen also provided feedback about pronunciation.

For accuracy in using vocabulary when speaking, the students' answers revealed improvement because the AI voice chatbot often suggested new vocabulary with an explanation while talking with the students. As a result, they were exposed to additional vocabulary.

The following responses give examples of the students' opinions regarding improving speaking accuracy in terms of grammar, pronunciation, and vocabulary.

> **PAR19:** *"I found that I could speak with more correct grammar and learn more structures from the AI voice chatbot because it used grammatical-corrected sentences when chatting with me. In addition, I could look at the chat contents displayed on the chat screen to find out how correctly I used grammar so that I could correct it by myself."*
>
> **PAR04:** *"When I pronounced wrongly, the chat screen displayed the wrong word. When I pronounced correctly, the chat screen displayed the right word. Therefore, I could look at that to self-evaluate my pronunciation. One more, the AI voice chatbot spoke with native voice, which supported me in learning pronunciation."*
>
> **PAR03:** *"The AI voice chatbot suggested more vocabulary related to our conversation whenever speaking, which inspired me to learn the words to be able to chat with it."*

### *C. The Improvement in Speaking Fluency*

Most of the participants said that they could speak with less hesitations compared practicing speaking with the AI voice chatbot. They explained that the first time, they spoke with hesitations and pauses. This prevented the AI voice chatbot from recognizing what they said. As a result, they had to practice speaking again and again until they could speak more fluently so that the AI voice chatbot properly recognize their speech. Students also reported that they felt free to speak with the AI voice chatbot and liked having unlimited time to speak repeatedly to help find speaking ideas. This reduced embarrassment when speaking with a real person. The reduced stress reduced their hesitations and pauses.

> **PAR03:** *"I can speak with less pauses compared with I did before. I could also speak more smoothly because I had time to find ideas. The AI voice chatbot could wait for me for a long time without making me embarrassed. I could also speak again and again until I got familiar with the speech flow gradually."*

**PAR05:** *"The AI voice chatbot did not limit the speaking time, so I could speak again and again many times, which helped me speak with less pauses and hesitations."*

***D. Positive Aspects of the AI Voice Chatbot***

When the participants were asked what they liked most about speaking with the AI voice chatbot, they reported that they liked to speak with the AI voice chatbot for a number of reasons. First, they said that the AI voice chatbot made them feel comfortable and gave them the sense of chatting with a human, which was inspiring. Second, they liked to speak with the AI voice chatbot because it pronounced accurately, like a native speaker which helped them improve their speaking skill in terms of pronunciation, grammar, and vocabulary without the embarrassment of speaking to a human. Participants said that the AI voice chatbot could speak for long time periods, which gave them more experience to practice English speaking.

**PAR03:** *"I found that the AI voice chatbot spoke very friendly and it could wait for me to repeat again and again. I could also speak with it at any time and any places. It could also speak like a native speaker and give me time to find ideas. Speaking with a real foreigner usually made me embarrassed with my incorrect pronunciation and long time for finding ideas."*

**PAR06:** *"Whenever we spoke, the AI voice chatbot frequently spoke, asked, and changed topics to keep the communication unstoppable."*

***E. Negative Aspects of the AI Voice Chatbot***

When asked about what they disliked, the students reported several reasons. First, the AI voice chatbot could not accurately understand the names of some places and names in Vietnamese. Second, the topics were still limited for students wanted more range in the content of practice conversations. Some did not like the app's frequent suggestions for buying premium features and its lack of in-app microphone.

**PAR03:** "T*he AI voice chatbot was still limited in terms of speaking topics. I wanted to speak about more and more topics but it could not support. One more thing is that it could*

*not recognize the name of Vietnamese places. Therefore, it usually displayed wrong names of such places."*

**PAR06:** "*What I dislike is that the app did not have integrated micro. I had to use the micro icon on my phone keyboard to speak. It would be great if a micro had been already integrated into the app directly."*

## 4.2 Discussion

The results of the study suggested that the students in both groups improved in English speaking skills. However, those in the experimental group obtained higher scores than those in the control group. Furthermore, the students in the experimental group improved in English speaking skills in terms of fluency and accuracy after they learned English speaking with an AI voice chatbot. Finally, they also agreed that using the AI voice chatbot could help them improve English speaking skills in terms of fluency and accuracy.

### 4.2.1 The Effects of an AI Voice Chatbot on English Speaking Skill of Vietnamese Undergraduate EFL Students

In discussing the effects of the AI voice chatbot on the students English speaking skill, two areas will be discussed, which include improvement in speaking accuracy and fluency, and differences in English Speaking Skills of the Students in the control group and in the experimental group.

#### 4.2.1.1 The Students 'Improvement in Speaking Accuracy and Fluency

The students speaking accuracy and fluency were improved, as demonstrated by improvement in scores of speaking accuracy and speaking fluency in the post-test, evaluated by the examiners using the speaking assessment rubric mentioned in Section 3.4.2.3. The reasons for the students improvement relate to the functions of the AI voice chatbot. First, the AI voice chatbot could correct students grammar when they spoke. The AI voice chatbot could inform the students of their mistakes in using quantifiers for uncountable and countable nouns and in differentiating between different verb tenses. This finding is quite a new discovery because previous studies (Fryer and Carpenter, 2006 Kim, 2016; Kim, 2017; Kim, 2018; Ahmad et al, 2018; Colace et al., 2018; Adamopoulou and Moussiades, 2020) did not clearly mention this function of the AI voice chatbots. Second, the AI voice chatbot could suggest vocabulary to the students to increase vocabulary when they practiced speaking. Within this function, the AI voice chatbot suggested words

randomly throughout the chat session along with definition in English so that the students could learn more vocabulary while speaking. In this way, the students knowledge of vocabulary was improved after they learned new words through conversation with AI voice chatbot. This finding provides a tool to help the students overcome their difficulties in achieving English speaking skill caused by insufficient vocabulary, mentioned by Dao (2017), Quyen et al, (2018), Tuyen and Loan (2019), Thao and Nguyet (2019). Finally, the AI voice chatbot had the ability to maintain the conversation with the students as it could continuously ask questions, respond to the students statements, and change speaking topics so that the conversation flow was always available. This finding is in line with Ahmad et al. (2018) that AI voice chatbots could speak with humans in answering and asking questions for extended periods of time.

**4.2.1.2 The Differences in English Speaking Skill of the Students in the Control Group and in the Experimental Group**

In investigating the students speaking scores in the pre-test, the scores were homogeneous between the control group and the experimental group. This finding agreed with Tuyen and Loan s (2019) suggestion that the students needed to be assigned to classrooms where the English level was at a similar level to provide ease and comfort in speaking. Moreover, the students scores were low in both groups, which was in line with Dao (2017) who noted that speaking skills were very difficult for students, and thus, they often received low scores in speaking tests.

The scores of the students in the experimental group, significantly improved on the post-test after they had practiced English speaking with the Andy English Bot app on their smartphone in the class. This finding suggests that using an AI voice chatbot can help the students improve their English speaking skills. This improvement came from the students increased practice time in actual English speaking during and after class thorough five weeks of experiment. This practice time met with Cambridge s (2017) suggestion that 25% of class time should be spent for practicing speaking to help students improve. In addition, the AI voice chatbot possessed a native voice which made the students feel as if they were speaking with a native speaker, and it brought about a stress-free speaking environment as they could speak freely and have few time limitations, both of which encouraged the students to speak more and thereby become more confident in speaking. Normally, students were nervous speaking with native speakers because they were afraid of spending too much time to find ideas to discuss, which might cause the discomfort

in the native speaker. However, when speaking with the AI voice chatbot, those difficulties were overcome.

This finding is in line with Nghi et al. (2019), that an AI voice chatbot could provide the students with new learning experiences which could improve the students learning performance when combined with class meetings. This was the reason why the students in the experimental group tended to demonstrate improvement in their speaking scores significantly after the intervention. Moreover, the finding is also in line with Kim s (2017) that using an AI voice chatbot could help the students improve their English speaking skills. In addition, when the students learned and practiced English speaking with Andy English Bot app on the smartphone, they were instructed carefully by a learning website where videos, images, and sounds were integrated. The students could also record their voice for self-assessment regularly. Therefore, this study also confirms the suggestions of Thao and Nguyet (2019) and Anh and Nhu (2021) that using visual aids could help the students speak with less stress, and that recording the students voices for later assessment could help them to be more self-aware in practicing English speaking.

Finally, the results of the improvement in speaking scores revealed that the students in the experimental group had higher scores than those in the control group in the post-test after the ten-week intervention. This finding suggests that using an AI voice chatbot or learning English speaking skills in a traditional classroom can improve the students speaking skills. However, using an AI voice chatbot can improve the students speaking skill more than learning English speaking skill in a traditional classroom because the AI voice chatbot could talk with the students at anytime and anywhere. This reduced barriers to practice such that students found it both easier and more motivating to speak when compared with speaking in traditional classrooms. This finding also agrees with Fulton s (2019) recommendation that using AI technologies like AI voice chatbots could provide the students with more chances to have English exposure to improve the students proficiency in the English language and help them to learn English more successfully.

**4.2.2 The Students 'Opinions on Using an AI Voice Chatbot**

Having investigated the students practice process with the students speaking scores in the post-test, the study found that those who spent their time speaking with Andy English Bot app on their smartphone for 10 minutes to 15 minutes a week got higher speaking scores than those who spent less than 10 minutes speaking per week. For example, PAR05 who spent 15 minutes speaking with the AI voice chatbot obtained 17.7 in the post-test while PAR10 who obtained 4.7 in

the post-test spent 5 minutes speaking with the AI voice chatbot every week. This can be explained by referring Cambridge s (2017) recommendation that 25% of the classroom time spent for practice is sufficient for the students to improve English speaking skills. Since the students learned English speaking for 135 minutes (three periods) a week, which took 25% of the total time of 540 minutes (twelve periods) for learning other English skills, this amount of time was sufficient for the students to improve their English speaking skills in general. However, if the students spent more time to practice English speaking with the AI voice chatbot after class, their improvement was more significant. This is why the students who spent more time practicing English speaking after class tended to get higher scores in the post-speaking test.

For the students opinions on improving English speaking accuracy, the study found that most of the students could speak with correct grammar, vocabulary, and pronunciation because they could self-evaluate their speaking performance based on the chat screen and audio recordings. Students reported that when they could speak and their utterances on the chat screen, it helped them to know if they had spoken with correct grammar and vocabulary. Moreover, they could also look at their utterances on the chat screen to see if they had just pronounced a word correctly because it would provide immediate feedback. In addition, the students self-study and self-evaluation abilities were reflected in their responses to the interview questions as they reported that when they listened to the AI voice chatbot s responses, they found that the AI voice chatbot used a variety of grammar and vocabulary from which they could learn more grammar and vocabulary simply by chatting with the AI voice chatbot. The students self-practice ability also enhanced pronunciation with practice. The AI voice chatbot could not understand what they said with poor pronunciation, which led them to improve. This finding is in line with the observations of Dao (2017), Quyen et al. (2018), Duc (2017), Tuyen and Loan (2019), and Thao and Nguyet (2019) who found that fostering the students self-awareness in learning enhanced their abilities of self-learning, self-practicing, self-evaluating, and self-creating.

In investigating the students opinions about their improvement in English speaking fluency, the study found that all students could speak better with less pauses and hesitations after they spoke with the AI voice chatbot during the 8-week intervention. This finding was demonstrated by students' reports that they could not speak too fast if they wanted the AI voice chatbot to understand them because the faster they spoke the more wrong pronunciations they made, reflected by the wrong

words appeared on the chat screen. Therefore, they had to speak carefully and repeatedly until the AI voice chatbot could recognize their words and understand what they had just said. The students said that they could speak again and again but the AI voice chatbot did not feel bored. It responded to the students at any time. This triggered a positive finding that the AI voice chatbot can be an attentive listener and speaker for the students to practice English speaking. In terms of theory, this finding agrees with Fryer and Carpenter (2006) and Ahmad et al. (2018) that the AI voice chatbot could tirelessly and attentively talk with students and be at their service all day. Moreover, the students reported that they felt like they were talking with a native speaker during conversations with the AI voice chatbot as it could speak smoothly with excellent pronunciation. This highlighted the characteristic of the AI voice chatbot of its ability to perform human-like English conversation as a native speaker, which is in line with Kim et al. (2021) and Çakmak (2022) that the AI voice chatbot had the ability to provide students with authentic inputs by speaking like a native speaker.

In exploring the students  feelings of using the AI voice chatbot to learn English speaking, the study found that almost all students agreed that they could improve English speaking skill after practicing speaking with the chatbot because they they felt relaxed when speaking with the AI voice chatbot as it was very friendly and patient. This finding agreed with the findings of Nghi et al. (2019), Kim et al. (2021), and Çakmak (2022) that an AI voice chatbot could bring relaxing and interesting learning environments to students to help them learn speaking more successfully and motivate them to learn and share their knowledge. However, there were some negative aspects of the AI voice chatbot reported by students such as the AI voice chatbot s  failure in recognizing names of some places in Vietnam, its  frequent popup advertisement, and its limited topic choice, although these drawbacks did not affect the students  English speaking skill.

## 4.3 Summary

In summary, this section has reviewed the results and of experiment and found crucial answers for the three research questions. The pedagogical implications provide useful clues for teachers and students to apply the AI voice chatbot to teach and learn English speaking skill successfully.

# CHAPTER 5
# CONCLUSION AND RECOMMENDATIONS

This research provided findings that the AI voice chatbot is effective in improving English speaking skills of Vietnamese undergraduate EFL students. There was a significant difference in the English speaking skill between Vietnamese undergraduate EFL students who study in the traditional classroom and those who study by using an AI voice chatbot. The positive opinions of the students toward using the AI voice chatbot for learning English speaking were discussed in Chapter 4. This chapter will summarize the major findings, and suggest implications of the study, in addition to considering the study limitations and provide recommendations for future study.

## 5.1 Summary of the Major Findings

The first finding of the study refers to the significant differences between the students English speaking skill in traditional classroom in comparison with the classroom learning English speaking skill with an AI voice chatbot. The study found that the students in both the control group and in the experimental group improved English speaking skills at the post-test. However, the students who were in the experimental group had a tendency to speak English better than those in the control group, as demonstrated by their English speaking post-test scores.

Second, when investigating the effects of an AI voice chatbot on English speaking skills of the students, this study found that the AI voice chatbot had useful functions for students to learn English speaking. It could help the students correct grammar mistakes, suggest vocabulary to students when speaking, and maintain the conversation flow with the students by asking follow-up questions or changing to a new topic to speak. Consequently, the longer time the student practiced English speaking with the AI voice chatbot, the higher speaking score they obtained in speaking fluency and accuracy, as demonstrated scores in their post-test. Their improvement was related to self-learning regarding grammar mistakes, updating their vocabulary knowledge, and speaking with the AI voice chatbot continuously.

The students opinions on using the AI voice chatbot to learn English speaking skill were generally positive because they found that they could improve their grammar, pronunciation, vocabulary, and fluency after speaking with the AI voice chatbot. Furthermore, the AI voice chatbot was also liked by the students because it

created a friendly and stress-free atmosphere while speaking with them like a native speaker.

## 5.2 Implications of the Study

The study s findings suggest that the AI voice chatbot has improved students speaking skills in terms of accuracy and fluency. It has also given the students more English speaking opportunities. More importantly, the results and the findings of this study suggest educational implications for students, pedagogical implications, and research implications.

### 5.2.1 Educational Implications for Students

For learning purposes, it is recommended that the students learn English speaking with others who are at an equivalent English level so that they can practice speaking easily with each other. This can also improve their confidence in speaking because it will lower the risk of feeling inferior to students.

When learning English speaking skill, students should use the AI voice chatbot to practice English speaking because the AI voice chatbot can help them speak and to self-assess their grammar, vocabulary, pronunciation, and pauses and hesitations. This self-awareness enables and encourages them to practice, evaluate, and learn more.

In addition, the students should also make use of the user interface or the chat screen to learn English speaking because the AI voice chatbot not only speaks with the students with a native voice, it can also display the chat contents as text messages which visually supports the students in both self-learning and self-practicing.

As the study found that the more time students spent practicing English speaking, the higher scores they obtained in the post-test, this suggests that when learning English speaking, students need to spend as much time as possible in practicing English speaking.

### 5.2.2 Pedagogical Implications

For pedagogical purposes, the first finding that teachers should be aware of is that they should assure equivalence of students' skills when they join an English class so that teachers can support students, and students can support other students.

As this study found that the students' English speaking skill improved more when they practiced English speaking with an AI voice chatbot, teachers are

encouraged to consider allowing students to use the AI voice chatbots in English speaking classrooms to practice English in addition to practicing with partners.

Sufficient time for practice should also be considered by teachers so that the practice time can be reasonably balanced. This means that time for practicing English speaking with the AI voice chatbot should be balanced with time for practicing with a partner. Teachers should not use only one method of practicing in such classrooms, because this would limit students chance to speak English.

Although the AI voice chatbot was found to improve the students accuracy and fluency in English speaking, it is important that teachers understand its use and its basic technology before introducing it into a classroom setting. Students need to be instructed and trained carefully before using the AI voice chatbot so that any technological problems can be avoided.

Regarding the AI voice chatbot s features, this study recommends that teachers take advantages of the AI voice chatbot s features such as text and voice support, unlimited conversation time, native voice, and attentive listener to give students more chances to speak and practice grammar, vocabulary, pronunciation. This allows them to improve their speaking accuracy first, after which they can speak with fluency. For example, when teachers teach a grammar point to students, teachers can ask students to practice a speaking exercise with the AI voice chatbot. After students finish speaking, they should then look at the chat screen to confirm if they have used correct grammar or not. For vocabulary and pronunciation, the process is similar.

**5.2.3 Research Implications**

For future research opportunities, this study first notes that AI voice chatbot is still a new field of TELL research, and therefore, the number of publications on AI voice chatbots remain limited. There will need to be many more future studies to provide additional data in this growing field.

Future research may delve into different aspects such as investigating how to measure students speaking fluency and accuracy when practicing English speaking with an AI voice chatbot, exploring how an AI voice chatbot helps students to practice language skills, exploring students attitudes toward using an AI voice chatbot to learn English speaking skills or other skills, or investigating how effectively an AI voice chatbot can improve students English proficiency in speaking, listening, writing, and reading. However, there are three things future researcher should consider.

The first is about technology. As AI voice chatbot is new in the educational area, technological problems will challenge both researchers and participants. Therefore, future researchers actively review and test various AI voice chatbots before selecting the one for their research. Within the scope of this study, Andy English Speaking Bot is a good suggestion. However, there are still many more AI voice chatbots which may work better such as ChatGPT, Nova AI voice chatbots, or GPT Chatvoice – AI Open Wisdom, although these chatbots require users to purchase access for use in research.

The second consideration future researchers should be related to the timeframe of the study. Ideally, a longer time period should be used for future experiments because 8 weeks may be inadequate to collect data on a newer technology like the AI voice chatbot.

Finally, future researchers should investigate the effects of the AI voice chatbot on other English language skills of students, especially in receptive skills like reading or listening. As Brown (2000) suggested, listening and speaking are two intertwined skills learners acquire when learning a language. It is strongly suggested that future researchers consider investigating the effects of the AI voice chatbot on English listening skills of students. This will bring about new vistas of exploration to this field of research.

## 5.3 Limitations and Recommendations for Future Studies

During the experimental period, there were some limitations which this study could not avoid. First, the experiment was conducted in eight weeks, which may have been too short a time for observing long-term development of the students. Further research should consider a longer timeframe to measure results. Second, this study had a small sample size, with 30 subjects in each arm. This may have limited some of the statistical analysis. Future researchers should consider studying a larger sample size to improve research data. Finally, technological issues were a challenge during the study because using an AI voice chatbot for learning English speaking skill is a very new field in Vietnam. Most of the English classrooms in Vietnam have been conducted traditionally. Therefore, both teachers and learners may lack familiarity with this new technology. As a result, many students had difficulties when using the AI voice chatbot for learning English speaking such as installing the app, activating the keyboard microphone to have a voice chat with the AI voice chatbot, and failing to activate the English voice so that the AI voice chatbot could recognize English words.

Fortunately, those particular problems were found during the orientation period. Hence, the researcher could help the students to solve each problem before they started using the AI voice chatbot for learning English speaking during the experiment.

Another technological issue was found but this issue could not be solved. It was about the inability of the AI voice chatbot to recognize Vietnamese names. Most of the time the AI voice chatbot could recognize Vietnamese names of locations in Vietnam such as Hue, Cham Temple, Hau Giang, Cai Rang. Regarding this issue, this author would like to suggest that future researchers pay careful attention to the technical aspects of the AI voice chatbot when conducting a study on using AI voice chatbots for learning other English skills.

## 5.4 Summary

This study investigated the effects of an artificial intelligence voice chatbot on English speaking skill of Vietnamese undergraduate EFL students. This study compared the significant differences in the English speaking skills between Vietnamese undergraduate EFL students who studied in the traditional classroom and those who studied by using an AI voice chatbot. Students opinions on using AI voice chatbots to improve English speaking skills have been also explored.

The results revealed that students improved their English speaking after learning English speaking and practicing English speaking with the AI voice chatbots in comparison to English speaking skills in a traditional classroom.

From the questionnaire and semi-structured-structured interviews, students also reported positive opinions on using the AI voice chatbots in English speaking to improve speaking fluency and accuracy in terms of grammar, vocabulary, pronunciation, pauses and hesitations.

For pedagogical applications, the AI voice chatbots are considered a good learning tool for the Vietnamese undergraduate EFL students as it is one of the technology-enhanced language learning methods. It also has support of the government and institutions. If teachers are ready and familiar with an AI voice chatbot, they can enhance interest and fun in classroom environments and engage students in learning English speaking skills. For students, the use of this technology provides a comfortable learning environment to practice English speaking and enhance English speaking skills.