

THE DEVELOPMENT OF LISTENING AND SPEAKING SKILLS IN EFL VIA
AN ARTIFICIALLY INTELLIGENT CHATBOT APPLICATION: A QUASI-
EXPERIMENTAL DESIGN STUDY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

FATMA ŞEYMA KOÇ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF ENGLISH LANGUAGE TEACHING

JANUARY 2024

Approval of the thesis:

**THE DEVELOPMENT OF LISTENING AND SPEAKING SKILLS IN EFL
VIA AN ARTIFICIALLY INTELLIGENT CHATBOT APPLICATION: A
QUASI-EXPERIMENTAL DESIGN STUDY**

submitted by **FATMA ŞEYMA KOÇ** in partial fulfillment of the requirements for
the degree of **Doctor of Philosophy in English Language Teaching, the Graduate
School of Social Sciences of Middle East Technical University** by,

Prof. Dr. Sadettin KİRAZCI
Dean
Graduate School of Social Sciences

Prof. Dr. Nurten BİRLİK
Head of Department
Department of Foreign Language Education

Prof. Dr. Perihan SAVAŞ
Supervisor
Department of Foreign Language Education

Examining Committee Members:

Assist. Prof. Dr. Müge GÜNDÜZ (Head of the Examining Committee)
Middle East Technical University
Department of Foreign Language Education

Prof. Dr. Perihan SAVAŞ (Supervisor)
Middle East Technical University
Department of Foreign Language Education

Prof. Dr. Çiler HATİPOĞLU
Middle East Technical University
Department of Foreign Language Education

Assoc. Prof. Dr. Servet ÇELİK
Trabzon University
Department of Foreign Language Education

Assist. Prof. Dr. Gamze ERDEM COŞGUN
Amasya University
Department of Foreign Language Education



I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Fatma Şeyma KOÇ

Signature:

ABSTRACT

THE DEVELOPMENT OF LISTENING AND SPEAKING SKILLS IN EFL VIA AN ARTIFICIALLY INTELLIGENT CHATBOT APPLICATION: A QUASI-EXPERIMENTAL DESIGN STUDY

KOÇ, Fatma Şeyma

Ph.D., The Department of English Language Teaching

Supervisor: Prof. Dr. Perihan SAVAS

January 2024, 319 pages

This research aimed to explore the effects of using an artificial intelligence chatbot to enhance the listening and speaking skills of Turkish EFL learners. It was also intended to specify the experiences of language learners and strategies for using this tool and identify the characteristics of an ideal artificially intelligent chatbot for language learning. Quantitative and qualitative data were collected towards these aims using a quasi-experimental design. The quantitative data were collected through a quasi-experimental pre- and post-test control group design. The qualitative data were collected through reflective diaries, recordings of student-chatbot interaction, and semi-structured interviews. According to the findings, although there were significant differences within groups regarding pre-test and post-test speaking proficiency exam results, there were no significant differences between groups regarding listening and speaking post-test scores. The qualitative analysis also demonstrated that although learners faced various technical, pedagogic, and social constraints, there were also self-reports of enhanced language knowledge and skills. The interaction data analysis also showed that EFL learners adopted several interactional strategies to negotiate for meaning in the cases of conversational breakdowns. Based on the findings of this study, a framework for developing artificial intelligent chatbots for EFL learners was

proposed. For this purpose, this study suggests a conceptual REPOST model, the guidelines for EFL teachers based on optimum AI chatbot qualities, and a teacher training component.

Keywords: MALL, AI Chatbot, English as a Foreign Language



ÖZ

İNGİLİZCE’NİN YABANCI DİL OLARAK ÖĞRENİMİNDE DİNLEME VE KONUŞMA BECERİLERİNİN YAPAY ZEKÂYLA GELİŞTİRİLMİŞ BİR SOHBET ROBOTU İLE GELİŞTİRİLMESİ: BİR YARI-DENEYSEL DESEN ARAŞTIRMASI

KOÇ, Fatma Şeyma

Doktora, İngiliz Dili Öğretimi Bölümü

Tez Yöneticisi: Prof. Dr. Perihan SAVAŞ

Ocak 2024, 319 sayfa

Bu araştırma, İngilizceyi yabancı dil olarak öğrenen Türk öğrencilerin dinleme ve konuşma becerilerini geliştirmede yapay zekâ chatbot kullanımının etkilerini araştırmayı amaçladı. Ayrıca dil öğrenenlerin deneyimlerini ve bu aracı kullanma stratejilerini belirlemek ve dil öğrenimi için ideal bir yapay zekâlı bir sohbet robotunun özelliklerini belirlemek de amaçlandı. Bu amaçlar doğrultusunda yarı deneysel bir araştırma deseni kullanılarak hem nitel hem de nicel veriler toplandı. Araştırmanın nicel verileri yarı deneysel ön-test son-test kontrol grubu yöntemiyle toplanırken, nitel veriler ise yansıtıcı günlükler, konuşma verileri ve görüşmeler yoluyla toplandı. Araştırmanın sonuçlarına göre her iki grup içerisinde konuşma testi ön test ve son test başarı puanları arasında anlamlı farklar bulunurken, gruplar arasında dinleme ve konuşma testi puanları arasında anlamlı bir fark bulunamamıştır. Nitel bulgulara göre ise katılımcılar teknik, pedagojik ve sosyal alanlarda güçlüklerle karşılaşmışlar, ancak bunun yanı sıra öz değerlendirmelere göre gelişmiş dil bilgi ve becerileri rapor edilmiştir. Konuşma verilerinin analiz sonuçları ise dil öğrenenlerin konuşma boyunca yaşadıkları güçlükler karşısında birçok strateji kullanarak iletişimi sağladıklarını

göstermiştir. Bu çalışmanın sonuçlarına bağlı olarak İngilizceyi yabancı dil öğrenenlere yönelik yapay zekâ destekli chatbotlar geliştirilmesi doğrultusunda kavramsal bir çerçeve bir öne sürülmüştür. Bu amaçlar doğrultusunda, REPOST adı verilen bir model, İngilizce öğretmenlerine yönelik kullanım yönergeleri ve bu doğrultuda bir öğretmen eğitimi modülü önerilmiştir.

Anahtar Kelimeler: Mobil Öğrenme, Yapay Zekâ Destekli Chatbot, İngilizcenin Yabancı Dil Olarak Öğrenimi



ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Perihan SAVAŞ, for her invaluable advice, continuous support, and consultancy during my PhD study. She has always supported and encouraged me whenever I encountered a trouble spot or had any concerns during my research.

I would like to express my gratitude towards my thesis committee members, Assoc. Prof. Dr. Servet ÇELİK and Assist. Prof. Dr. Müge GÜNDÜZ, for their invaluable assistance and insightful comments and suggestions at every stage of my research. I would also like to extend my thanks to my thesis defense jury members, Prof. Dr. Çiler Hatipoğlu and Assist. Prof. Dr. Gamze ERDEM COŞGUN, for their invaluable advice on further developing the ideas in my thesis.

I would like to thank Assoc. Prof. Dr. Hüseyin KAFES for helping me during the data collection process by allowing me to approach his students as the research participants in this study. I also thank the research participants in this study for all the effort they have put into carrying out their assigned tasks.

I also thank Turkey's Scientific and Technological Research Council (TÜBİTAK) for financial support throughout my doctoral studies.

I am deeply grateful to my dear friend and colleague Özlem GÜNEŞ ŞENOĞLU for her support during the assessment phase of the speaking proficiency exams and the qualitative data analysis process. She has spent her valuable time contributing to my study.

I sincerely thank my friend and colleague, Dr. Funda ÖLMEZ ÇAĞLAR, for her friendship and support during this process.

I would also like to thank my brother Aykut Alparslan KOÇ for his help with the transcriptions of the recordings. He has developed a computer program that transcribes audio via an AI model and has done the preliminary transcriptions.

Finally, I must express my profound gratitude to my parents, Filiz KOÇ and Dursun KOÇ, for their endless support, care, and love.



TABLE OF CONTENTS

PLAGIARISM.....	iii
ABSTRACT.....	iv
ÖZ	vi
ACKNOWLEDGMENTS	viii
TABLE OF CONTENTS.....	x
LIST OF TABLES	xiv
LIST OF FIGURES	xvii
LIST OF ABBREVIATIONS.....	xviii

CHAPTERS

1. INTRODUCTION	1
1.0. Presentation	1
1.1. Background to the Study	1
1.2. Statement of the Problem	6
1.3. Purpose of the Study	8
1.4. Significance of the Study	10
2. LITERATURE REVIEW	12
2.0. Presentation	12
2.1. Intelligent Computer and Mobile Assisted Language Learning	12
2.2. Negotiation of Meaning	14
2.3. Studies in the Field of Language Teaching.....	19
2.3.1. Chatbots in EFL.....	21
2.3.2. Benefits of Chatbots in EFL	28
2.3.3. Challenges of the Use of Chatbots in EFL	31
3. METHODOLOGY	34
3.0. Presentation	34
3.1. Overview of the Research Design.....	35
3.2. A Quasi-Experimental Pretest-Posttest Control Group Design	36
3.3. Participants.....	40
3.4. Data Collection Instruments.....	41
3.4.1. Pre-Test and Post-Test.....	42

3.4.1.1. Listening Test.....	44
3.4.1.2. Speaking Test.....	45
3.4.1.3. Validity and Reliability of the Proficiency Tests.....	46
3.4.2. Semi-structured Interviews	47
3.4.3. Reflective Diaries.....	48
3.4.4. Credibility and Trustworthiness.....	48
3.4.5. AI Chatbot-English Language Learner Interaction Data	50
3.4.6. The AI Chatbot: Replika	50
3.5. Piloting.....	54
3.6. Data Collection Procedures.....	55
3.7. Treatment	57
3.8. Data Analysis	59
3.8.1. Quantitative Analysis: Statistical Procedures	61
3.8.2 Coding Procedures for Qualitative Analyses	62
3.8.3. The Analysis of the ELL-AI Chatbot Interaction Data.....	64
3.9. The Background of the Assessors of the Speaking Exam	65
3.10. The Role of the Researcher.....	65
3.11. Triangulation.....	66
4. FINDINGS.....	68
4.0. Presentation.....	68
4.1. Findings in Relation to the Comparison of the Experimental and the Control Groups on Listening Skills Development.....	68
4.1.1. Descriptive Statistics of the Listening Pre-tests.....	68
4.1.2. Descriptive Statistics of the Listening Post-tests	70
4.1.3. The Normality Tests of the Listening Scores.....	72
4.1.4. Comparison of the Listening Pre-test and the Post-test Scores.....	73
4.2. Findings in Relation to the Comparison of the Experimental and the Control Groups on Speaking Skills Development.....	75
4.2.1. Descriptive Statistics of the Speaking Pre-tests	75
4.2.2. Descriptive Statistics of the Speaking Post-tests	77
4.2.3. The Normality Tests of the Speaking Scores.....	79
4.2.4. Comparison of the Speaking Pre-test and the Post-test Scores.....	80
4.3. Findings in Relation to the Types of Negotiation of Meaning Sequences in AI Chatbot-student Communication.....	81
4.4. Findings in Relation to the Technological Affordances of Using AI Chatbot Technology for English Language Learning.....	93

4.4.1. Findings Regarding Technological Opportunities from the Reflective Diaries.....	97
4.4.2. Findings Regarding Technological Constraints from the Reflective Diaries.....	106
4.4.3. Findings Regarding Technological Opportunities from the Interviews.....	128
4.4.4. Findings Regarding Technological Constraints from the Interview Data.....	134
4.5. Findings in Relation to the Pedagogical Affordances of Using AI Chatbot Technology for English Language Learning.....	152
4.5.1. Findings Regarding Pedagogical Opportunities from the Reflective Diaries.....	154
4.5.2. Findings Regarding Pedagogical Constraints from the Reflective Diaries.....	158
4.5.3. Findings Regarding Pedagogical Opportunities from the Interview Data.....	160
4.5.4. Findings Regarding Pedagogical Constraints from the Interview Data.....	170
4.6. Findings in Relation to the Social Affordances of Using AI Chatbot Technology for English Language Learning.....	177
4.6.1. Findings in Relation to Social Opportunities from Reflective Diaries.....	179
4.6.2. Findings on Social Constraints from Reflective Diaries.....	184
4.6.3. Findings Regarding Social Opportunities from the Interview Data.....	188
4.6.4. Findings Regarding Social Constraints from the Interview Data.....	191
5. DISCUSSION.....	196
5.0. Presentation.....	196
5.1. Discussion in Relation to Comparison of the Experimental and the Control Groups on Listening Skills Development.....	196
5.2. Discussion in Relation to Comparison of the Experimental and the Control Groups on Speaking Skills Development.....	198
5.3. Discussion in Relation to the Types of Negotiation of Meaning Sequences in AI Chatbot-Student Communication	201
5.4. Discussion in Relation to the Technological Affordances of Using AI Chatbot Technology for English Language Learning.....	206
5.5. Discussion in Relation to the Pedagogical Affordances of Using AI Chatbot Technology for English Language Learning.....	212
5.6. Discussion in Relation to the Social Affordances of Using AI Chatbot Technology for English Language Learning.....	222

5.7. A Framework for AI Chatbot Use for EFL Teachers in Teaching Target Skills Based on the Findings	229
5.7.1. Optimum Features of AI-based Chatbots for Learning and Practicing English as a Foreign Language.....	230
5.7.2. Guidelines for Teachers Based on the Characteristics of Optimum AI Chatbot Features.....	239
5.7.3. Suggested Conceptual Model in Relation to Language Knowledge Attainment and Skills Practice	242
5.7.4. Suggestion for an Integrated Module on AI Chatbot-assisted English Language Learning.....	248
5.7.5. Implications for EFL Classes and Language Teachers	251
5.7.6. Implications for Application Developers	254
5.7.7. Implications for Teacher Educators	256
6. CONCLUSION.....	258
6.0. Introduction.....	258
6.1. Summary of the Study	258
6.2. Conclusion	259
6.3. Limitations and Suggestions for Further Research.....	261
REFERENCES	263
APPENDICES	284
A. DEMOGRAPHIC INFORMATION QUESTIONNAIRE.....	284
B. SEMI-STRUCTURED INTERVIEW QUESTIONS	286
C. REFLECTIVE DIARY PROMPT	287
D. APPROVAL OF THE METU HUMAN SUBJECTS ETHICS COMMITTEE.....	291
E. CURRICULUM VITAE	292
F. TURKISH SUMMARY / TÜRKÇE ÖZET.....	295
G. THESIS PERMISSION FORM / TEZ İZİN FORMU	319

LIST OF TABLES

Table 1 The design of the quasi-experiment.....	38
Table 2 Cambridge CAE Test.....	43
Table 3 Methodology outlined.....	59
Table 4 Descriptive statistics of the listening pre-test.....	69
Table 5 The proficiency levels of students on the listening pre-test.....	69
Table 6 Descriptive statistics of the listening post-test.....	70
Table 7 The proficiency levels of students on the listening post-test.....	71
Table 8 Normality tests for the listening pre-test.....	72
Table 9 Findings of normality tests for the listening post-test.....	72
Table 10 Mann Whitney U Test for listening pre-test scores among two groups.....	73
Table 11 Mann Whitney U Test for listening post-test scores between the two groups.....	74
Table 12 Paired Samples T-test results for the listening test scores of the experimental group.....	74
Table 13 Wilcoxon Signed Ranks Test results for the listening test scores of the control group.....	75
Table 14 Descriptive statistics of the speaking overall score pre-test.....	75
Table 15 The proficiency levels of students on the speaking pre-test.....	76
Table 16 Descriptive statistics of the speaking post-test.....	77
Table 17 The proficiency levels of students on the speaking post-test.....	77
Table 18 Skewness and kurtosis values of the experimental and control groups in speaking pre-test exam results.....	79
Table 19 Skewness and kurtosis values of the experimental and control groups in speaking post-test exam results.....	79
Table 20 Independent samples t-test results for speaking pre-test scores.....	80
Table 21 Independent samples t-test results for speaking post-test scores.....	80
Table 22 Paired samples t-test findings for the speaking scores of the experimental group.....	81

Table 23 Paired Samples T-test results for the speaking scores of the control group.....	81
Table 24 Amount and type of NfM stages in seven weeks.....	82
Table 25 Types of negotiation of meaning signals.....	82
Table 26 Types of triggers.....	83
Table 27 Response types.....	84
Table 28 Total quantity of modified output.....	84
Table 29 Types of modified output.....	85
Table 30 An example of negotiation of meaning routine by ELLs.....	85
Table 31 An example of negotiation of meaning routine by ELLs.....	86
Table 32 An example of negotiation of meaning routine by ELLs.....	87
Table 33 An example of negotiation of meaning routine by ELLs.....	88
Table 34 An example of negotiation of meaning routine by ELLs.....	89
Table 35 An example of negotiation of meaning routine by ELLs.....	89
Table 36 An example of negotiation of meaning routine by ELLs.....	90
Table 37 An example of negotiation of meaning routine by ELLs.....	91
Table 38 An example of negotiation of meaning routine by ELLs.....	92
Table 39 Overview of the frequencies of technological affordances codes from reflective diaries and interviews.....	93
Table 40 Frequencies for technological opportunities codes emerging from the reflective diary data.....	97
Table 41 Frequencies for technological constraints codes emerging from the reflective diary data.....	106
Table 42 Frequencies for technological opportunities codes emerging from the interview data.....	128
Table 43 Frequencies for technological constraints codes emerging from the interview data.....	135
Table 44 Frequencies for pedagogical affordances codes emerging from reflective diaries and interviews.....	152
Table 45 Frequencies for pedagogical opportunities codes emerging from the reflective diary data.....	154

Table 46 Frequencies for pedagogical constraints codes emerging from reflective diary data.....	158
Table 47 Frequencies for pedagogical opportunities codes emerging from interview data.....	161
Table 48 Frequencies for pedagogical constraints codes from the interview data.....	170
Table 49 Frequencies for social affordances codes emerging from reflective diary and interviews.....	178
Table 50 Frequencies for social opportunities codes emerging from reflective diary data.....	179
Table 51 Frequencies for social constraints codes from the reflective diary data.....	184
Table 52 Frequencies for social affordances codes from the interview data.....	188
Table 53 Frequencies for social constraints codes emerging from the interview data.....	191
Table 54 Optimum technological AI chatbot features based on the findings of this study.....	231
Table 55 Suggestions for implementation based on the optimum AI-enabled chatbot features.....	239
Table 56 Suggested module for training on AI-based chatbot creation for EFL teachers.....	249
Table 57 Suggested module for training on AI-based chatbot creation for EFL teachers.....	251

LIST OF FIGURES

Figure 1	The procedure for the quasi-experimental design in this study.....	38
Figure 2	Participant selection process.....	40
Figure 3	Qualitative data collection tool preparation steps.....	49
Figure 4	The mobile version of the AI Chatbot Replika.....	52
Figure 5	The desktop version of Replika.....	53
Figure 6	The intervention design of the study.....	58
Figure 7	The REPOST Model.....	244
Figure 8	Objectives for a workshop on AI chatbot creation for EFL teachers.....	248
Figure 9	Phases of integration of AI-based chatbots in language classes.....	252

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
API	Application Programming Interface
ASR	Automatic Speech Recognition
CALL	Computer-Assisted Language Learning
ChatGPT	Chat Generative Pre-Trained Transformer
CMC	Computer-Mediated Communication
EFL	English as a Foreign Language
ELL	English Language Learner
ELT	English Language Teaching
ESL	English as a Second Language
GA	Google Assistant
I-CALL	Intelligent Computer Assisted Language Learning
IPA	Intelligent Personal Assistant
L2	Second Language
MALL	Mobile Assisted Language Learning
M-learning	Mobile Learning
NfM	Negotiation for Meaning
NLP	Natural Language Processing
WTC	Willingness to Communicate
XR	Extended Reality

CHAPTER I

INTRODUCTION

1.0. Presentation

The first chapter of this dissertation study begins with a demonstration of the overview of the historical developments in artificial intelligence-powered mobile language learning systems. The types of AI chatbots created based on different knowledge systems are described in light of the developments in the latest technologies. The rationale behind the implementation of this study is explained in line with the needs of English language learners within the national context in which the research was conducted, as well as the findings of the studies carried out in the international literature. The contribution of this study to the field of English language teaching has been discussed through the exploration of the research questions constructed to meet the aims of this research.

1.1. Background to the Study

In the 1970s, communication through computer systems became possible with the invention of dialogue systems, and after 20 years, chatbots using text-based communication modes were created by programmers (J.H. Lee et al., 2020). Intelligent chatbots carry out tasks such as looking for information and playing songs; interaction through these devices is transactional (Dizon, 2021). Chatbots, called chatterbots, talkbots, or bots, are computer programs that can imitate an exchange with users and have the potential to enhance interaction opportunities through m-learning (Bailey, 2019; Troussas et al., 2017).

As defined by Pérez et al. (2020), “a chatbot is a tool that combines artificial intelligence (AI) and natural language processing and other technology, which enables

it to interact to a certain level of conversation with a human interlocutor through text or voice” (p.1). Using intelligent virtual assistants that are also named “human–computer dialogue systems, machine conversation systems, virtual agents, dialogue systems, chatterbots” (Jia & Chen, 2009, p. 903), digital assistants, and chatbots (Dale, 2016), communication can be carried out in a dialogic manner through natural language. Intelligent personal assistants are constructs with which users can communicate via text or audio, like Amazon’s *Alexa*, Apple’s *Siri*, *Cortana*, or *Google Assistant* (Ali Amer Jid Almahri et al., 2019). This technology aims to copy the neural networks of the human nervous system and mimic how humans remember, contemplate, and talk (Fryer, Coniam, Carpenter & Lăpuşneanu, 2020).

J.H. Lee et al. (2020) propose a classification system for these AI-based chatbots based on the categories of purpose and mode of communication. Regarding purpose, chatbots can be categorized into assistant and conversation types. Assistant chatbots like *Siri* and *Cortana* are programmed to help users in daily life with activities such as looking for information on Google, calling people, making appointments, etc., by integrating AI into mobile devices. The communication of these assistants is concise, systematic, and straightforward. On the other hand, conversation chatbots provide lengthier input by getting involved in human-like chats with their users.

A classification system for chatbots regarding speaker and technology dimensions has also been proposed. The speaker dimension includes “skill level, language diversity, domain, and interaction modality” (Petrović & Jovanović, 2021, p. 317). The skill level is about the chatbot’s adaptability to the language proficiency of speakers. Language diversity refers to being suitable for learners from various language backgrounds. The domain is about the chatbot’s capability to elaborate themes and subjects. In addition, the modality of the chatbot could be textual, voice-based, or both. Regarding the technology dimension, the chatbots could be categorized under “dialog type, knowledge base, and availability” (Petrović & Jovanović, 2021, p. 318).

The first component, dialogue type, could either be predefined, based on several rules, or statistical, which depends on machine learning algorithms from more extensive datasets. *Mitsuku* is one example of a statistical chatbot type with an AI markup

language system choosing its answers based on keyword selection from a knowledge database. On the other hand, the knowledge base refers to a system composed of linguistic concepts and phrases of the language. At the same time, the availability relates to the options of being free of charge or entirely or partly charged.

Thus, constructed based on various dialog management systems, intelligent personal assistants are utilized in many areas, such as education, healthcare, commerce, and entertainment (Ali Amer Jid Almahri et al., 2019). Chatbots have been used in various domains including but not limited to customer service (Chung et al., 2020; Cui et al., 2017), healthcare (Divya et al., 2018), digital psychological counseling (A. Ho et al., 2018), finance (Kuligowska, 2015), tourism (Ukpabi et al., 2019), sports (Segura et al., 2019), insurance (Nuruzzaman & Hussain, 2020), and traveling (Argal et al., 2018).

Supporting the view that the implementation of artificially intelligent I-CALL systems contributes to the pedagogical fields, AI chatbots are utilized within the education system to enable students to have self-guidance, develop their listening and comprehension skills, motivation, and enjoyment (Ali Amer Jid Almahri et al., 2019). Thus, chatbots are being specifically developed for education, and the purpose of designing chatbots in education is to produce knowledge just as a human instructor would. Consequently, the number of chatbots intended for particular groups of students on specific topics has increased (Pérez et al., 2020). In the field of education, chatbots have been implemented in the areas such as distance education (Heller et al., 2005), education support systems for frequently asked questions (Patel et al., 2019; Villegas-Ch et al., 2020), nursing education (Chang et al., 2020), medical education (Kaur et al., 2021), science education (Durall & Kapros, 2020; Topal et al., 2021), and in teaching languages such as Spanish language (Segura et al., 2019) and Chinese language (H.L. Chen, Vicki Widarso & Sutrisno, 2020).

Researchers have proposed various reasons for using AI chatbots in their teaching contexts. For instance, the property of speech recognition software and the widespread use of mobile devices have been indicated as factors that motivate their use in language learning (Fryer, Coniam, Carpenter & Lăpuşneanu, 2020). Artificial intelligence and

machine learning embedded in these application interfaces (APIs) also allow the chatbots to recognize unstructured user inputs (Huang et al., 2022), which might add to the strengths of these tools for the field of language learning in which the comprehensibility of the utterances by non-native users of languages might impact the effectiveness of the use of AI chatbots. Other researchers have argued that thanks to artificially intelligent chatbots, language learners can practice language skills almost anytime and anywhere, even without teachers' support. The studies on the use of AI chatbots have also explored this technology's contribution to improving English language knowledge and skills in speaking (M.H. Hsu, Chen & Yu, 2023; N.-Y. Kim, 2017; H.S. Kim, Kim, & Cha, 2021; C.J. Lin & Mubarak, 2021; Sydorenko et al., 2018; Sydorenko et al., 2019; Yin & Satar, 2020), listening and speaking (Dizon, 2020; Meri- Yilan, 2019), writing (M.P.C. Lin & Chang, 2020; S.H. Yang, 2007), and prepositions (Nghie et al., 2019) in English as a foreign language (EFL) and English as a second language (ESL) settings. Moreover, in the literature, some researchers developed chatbots for learning English (Goda et al., 2014; Jeon, 2022b; Johnson, 2019; Pham et al., 2018; Sung, 2020; Xu et al., 2021).

Using visual chatbot development avenues, it is now possible for teachers to develop chatbots even though they do not have any background in programming. One such platform is *Dialogflow* by Google (<https://dialogflow.com>), via which chatbots could be created using pre-set data to construct conversational scenarios. *BotStar* is another chatbot design avenue through which teachers could create conversation flows to match their students' learning needs. Developing chatbots by teachers could be a promising area for the enhancement of the opportunities these tools would provide to educational settings because, through the assessment of the functioning capabilities of chatbots on learners with actual classroom use data, chatbots meeting the profiles and needs of a specific group of learners could be created. It is thus possible to develop language models in AI chatbot systems based on the needs of a particular group of learners (Xu et al., 2021).

Besides the development of AI chatbots by the specialists in the field, the use of non-goal-aligned chatbots such as *Alexa* and *Siri* by teachers of foreign and second languages is also common, as they could be installed on mobile phones or desktop

computers equipped with auditory gears. Therefore, researchers also choose to implement already present publicly available AI chatbots like *Alexa*, *Siri*, and *Cortana* (Dizon, 2017; Moussalli & Cardoso, 2020), *CSIEC* (Jia, 2009), *Mitsuku* (Shin et al., 2021), *Cleverbot* (Thompson et al., 2018), *ALICE* (Troussas et al., 2017), *Replika* (Kılıçkaya, 2020; C.J. Lin & Mubarak, 2021). However, in contrast to the pedagogical AI chatbots developed by teachers for their teaching settings, these publicized chatbots must be adapted to the teachers' requirements of specific educational environments. In other words, as chatbots like *Alexa* and *Siri* are initially created for general audiences, teachers must adapt these tools according to their instructional aims. For instance, based on the technical characteristics of chatbots, teachers could provide language learners with specific commands to use the applications.

Thus, using mobile-assisted language learning (MALL) applications equipped with AI technologies in language learning involves using applications created explicitly for learning languages or general audiences. However, utilizing both types is considered influential thanks to their adaptability for language learning purposes (Kukulska-Hulme, 2020). Therefore, these non-goal-aligned AI chatbots targeting general audiences also function as interaction partners for the practice of the target language by language learners and would help learners interact in a dialogical manner. Consequently, considering the developments in technology and their recent applications in the field of education and language learning, it could be put forward that the fast advancement of computer-assisted language learning (CALL) compelled the education process to diverge from teacher and learner-focused communication and text-based format, and opened up new possibilities for captivating and compelling learning experiences (Elaish et al., 2017; Jeon, Lee & Choi, 2023).

Learners who continue to learn beyond the traditional or virtual classes with portable mobile devices could become lifelong learners who could improve their language competencies more independently due to abundant multimedia sources. Utilizing portable devices could also aid learners who may not feel confident enough to practice learning languages in public and would like to have a private learning sphere to practice their language skills, such as speaking and pronunciation, as even when they are at home, they may have less personal space to use a shared computer (Kukulska-

Hulme, 2020). Due to the possibilities AI chatbots promise to their users through being omniscient language learning and practice partners, these AI tools might even hold more prospects for the field of English language teaching in the future by compensating for the lack of interactional chances language learners in EFL or ESL contexts would need.

1.2. Statement of the Problem

Moving on to the rationale for conducting this study, among the reasons for challenges in EFL speaking classes in Turkey, a lack of second language (L2) input has been one prominent factor, and a lack of practice of language skills is often reported by learners (Kara et al., 2017). One of the issues surrounding English as a foreign language (EFL) settings is the need for more opportunities to practice English because English learners can only sometimes have enough language skills to practice with their peers or teachers in class. Besides, language learners have few chances for oral language skills practice outside the class in these EFL contexts. One way teachers could support language learners is the provision of technological tools adaptable for use in language learning.

Language teachers have been adopting computer-mediated means to alleviate the lack of chances to practice linguistic skills. With the recent advances in AI technology, intelligent computer-assisted language learning tools are being adopted to provide learners with the optimal conditions for practicing languages. As one intelligent language learning tool, an AI chatbot offers prospects for exchanging conversations with AI avatars in a human-resembling mode of communication. It is due to the current progressions in AI technologies that have now made it possible for computer systems to recognize the meaning of the statements uttered by speakers and respond to them correspondingly (C.J. Lin & Mubarak, 2021) and respectively act as conversation partners for language learners. Consequently, chatbots offer learners a way to interact more, especially with few opportunities to practice language outside the classroom (Fryer & Carpenter, 2006). Therefore, using artificial intelligence (AI) technologies raises the possibility for learners to practice learning English in and out of the classroom. Chatbots also carry the potential to turn the learning experience into a more exciting and meaningful journey by delivering an authentic mode of communication

(S.H. Yang, 2007). It is thus proposed that “chatbots are precisely the kind of educational technology that many instructors reach for when seeking to increase students' behavioral engagement” (Fryer, Thompson & Nakao et al., 2020, p.3).

In the literature, there are several studies conducted on the use of chatbots in language learning; however, the chatbots implemented in these studies have only a text-based chat function, which limits the communication to merely the written mode (Fryer et al., 2019; Fryer, Coniam, Carpenter & Lăpușneanu, 2020; Yin & Satar, 2020). Only in recent years has there been an increase in the use of AI chatbots with the integration of voice-based functionality due to the improvements in natural language processing (NLP), automatic speech recognition (ASR), and text-to-speech technologies. Studies implemented with these more advanced AI chatbots with voice recognition and artificial intelligence capabilities have shown that the use of AI chatbots positively impacts the speaking skills of English language learners (Dizon, 2020; El Shazly, 2020; H.S. Kim, Kim, & Cha, 2021; C.J. Lin & Mubarak, 2021), improve willingness to communicate levels (Ayedoun et al., 2019; Tai & Chen, 2020), L2 confidence (H.H.J. Chen, Yang & Lai, 2023; Jeon, 2022b; Tai & Chen, 2020) and interest in learning languages (Fryer et al., 2019), and lowers speaking anxiety (L. Hsu, 2022; Tai & Chen, 2020).

In recent years, there have also been attempts to create AI chatbots by specialists in the field specifically for the pedagogical purposes of learning and teaching the English language (Ayedoun et al., 2019; Bailey et al., 2021; Johnson, 2019). However, although these learning tools can provide extracurricular support for practicing English skills, especially in EFL settings, their use still needs to be improved.

Research is scarce in this field, and more specifically, more studies need to explore the use of AI chatbots in learning languages with a voice-based function in the Turkish EFL context. Researchers in Turkish EFL contexts have utilized chatbots primarily based on textual-mode communication (Kılıçkaya, 2020; Meri-Yılan, 2019). For these reasons, future studies that implement AI chatbot technology in language teaching would help foster autonomous learning, support extracurricular activities for learning

languages, increase opportunities for language practice through oral communication, develop language skills, and decrease anxiety in learning languages.

The researchers also need to design AI chatbots suitable for a specific group of target language learners. More specifically, the number of publicly available voice-enabled AI chatbots created for teaching and practicing languages must be increased. Language learning applications available for learners on the application stores of mobile devices also allow their users to communicate based on a rote learning mode. For instance, in *Duolingo*, the users can practice speaking skills in various languages by selecting pre-set options in speaking scenarios and reading them aloud. Consequently, language learners cannot interact through open-ended dialogues in a two-way communication. Instead, the speaking exercises in these applications are based on close-ended interactions.

Moreover, regarding improving speaking skills, the currently available pedagogical language learning applications also provide intonation and pronunciation practice exercises in which learners must read some texts aloud. The system provides users with input by providing a percentage of the words recognized by its system. It gives overall speaking scores to its users on the aspects of pronunciation and intonation. However, this system is again based only on automatic speech recognition system feedback, not having been developed based on the components of large language models, natural language processing, and artificial intelligence. As a result, considering the potential of these tools for language learning and practice, there is a need to create chatbots best suited to the preferences and needs of language learners. The experiences of students reported would help demonstrate the characteristics that this generation of AI chatbots offers in language learning and the deficiencies that impede the interaction opportunities these agents provide.

1.3. Purpose of the Study

This study aims to demonstrate if using an artificially intelligent chatbot mobile application as an extracurricular activity would contribute to developing listening and speaking skills in the English language. With the employment of an experimental

methodology, the impact of the regular use of this novel technology in learning a foreign language as supplementary material is observed, and the findings of this study set the ground for more research to be conducted regarding the use of AI technologies in language learning. In addition to the experimental quantitative measures, qualitative data collection methods aim to gain the language learners' perspectives on the beneficial and unfavorable sides of learning languages with AI technology. The findings of this study would highlight the features of effective AI chatbots for learning foreign languages. Based on the findings of this study, practical features that are expected to be possessed in an AI chatbot application will be demonstrated through the representation of English language learners' perspectives. Analyzing the AI chatbot-English language learner interaction data would also depict the interactional opportunities these tools provide in language classes. The AI-based interaction analysis would help depict how these tools function during a conversation with EFL learners. Therefore, the following research questions were generated for this study in line with the related conceptual frameworks and literature.

1. Are there any significant differences between the control group and the experimental group engaged in interacting with an artificially intelligent chatbot among Turkish tertiary EFL learners regarding
 - a. listening comprehension scores?
 - b. speaking scores?
2. Are there any significant differences in the pre-test and post-test language proficiency test scores within the control and experimental groups after the treatment period regarding
 - a. listening comprehension scores?
 - b. speaking scores?
3. What strategies are used to negotiate meaning in the AI chatbot-student communication interaction data?
 - a. Which stages of NfM form the NfM routines in Turkish EFL learners' interactions with the AI chatbot *Replika*?

- b. Which types of triggers lead to communication breakdowns in the NfM routines?
 - c. What is the number of modified output instances produced by Turkish EFL learners' during their interactions with the AI chatbot *Replika*?
- 4. Based on the perceptions of the participants of the study:
 - a. What are the strengths of using chatbot technology for English language learning?
 - b. What are the weaknesses of using chatbot technology for English language learning?
 - c. What are the features of an ideal AI chatbot for English language learning?

1.4. Significance of the Study

AI chatbots can be essential means to motivate students and elevate interaction. They can be utilized according to different individual characteristics and needs. Intelligent bots could also decrease feelings of solitude (Topal et al., 2021). Learners can be motivated by this mode of communication, as they can perform hands-on activities through voice commands. Moreover, they have the potential to capture students' interest. They also assist in cases in which teachers are unable to give answers to students on topics they are not familiar with. They also support teachers during class time when they cannot attend to each student in class by enabling students to carry out group work and peer work activities (Underwood, 2021).

However, despite these contributions, the research on implementing artificially intelligent chatbots in English as a foreign language is limited. A few chatbots are also created for language users to help them interact via voice recognition or multimodal means. More specifically, there is a need to study further the impacts of AI chatbots on the development of speaking, listening, reading, and writing skills (N.Y. Kim, Cha & Kim, 2019). Although there is a vast body of research on MALL and its impacts on L2 writing, the research on MALL and L2 speaking is restricted. However, thanks to

advances in NLP and ASR technologies, research in dialogue-based CALL adopting interactionist theories has become more widespread (Dizon, 2020). Thus, this study would demonstrate the impacts of using an AI chatbot on the development of Turkish EFL learners' listening and speaking skills and illuminate the characteristics expected to be inherent in these applications with the most up-to-date voice-enabled AI chatbot technologies.

Furthermore, as EFL students used the AI chatbot as extracurricular activities in this study, it aims to help foster learner autonomy when students self-regulated the time they spent with the AI chatbot. During the time frame in which this study was conducted, the students in this education context attended the classes through a hybrid teaching mode due to the restrictions caused by the COVID-19 pandemic. For this reason, the instructors delivered the lessons face-to-face and broadcasted them online. Therefore, using an AI chatbot companion for these students promoted interaction opportunities for listening and speaking skills practice by acting as speech partners for these EFL learners.

CHAPTER II

LITERATURE REVIEW

2.0. Presentation

The second chapter of this thesis provides a portrait of the applications of intelligent computer-assisted and mobile tools-supported language learning systems. The theoretical backgrounds of this study based on interactionist perspectives are highlighted. Furthermore, the methodologies adopted, research contexts, participants, and the specific types of AI-based chatbot tools implemented in the studies are described. The demonstration of the findings is constructed through a focus on the findings in the literature regarding the adoption of AI-based chatbots in EFL settings. The benefits and potential drawbacks of using these applications for language learning and skills practice are reported. The future implications of these studies are reviewed based on the reported language learning outcomes and the self-reports of language learners, teachers, and educational technology specialists.

2.1. Intelligent Computer and Mobile Assisted Language Learning

The recent research focus on Intelligent Computer Assisted Language Learning (ICALL), which surrounds the concepts of “computational linguistics, language engineering, natural language processing, and artificial intelligence” integrated with CALL systems (Dodigovic, 2005, p.98) has demonstrated a shift towards the inclusion of these Natural Language Processing (NLP) facilities and appliances into CALL or MALL devices that allow processes such as analysis of morphological and semantic features, tagging of parts of speech, and utilizing tools such as tokenizers, chunkers, and parsers that control the syntax and semantics of utterances (Amaral & Meurers, 2011). The use of mobile tools in learning could be elevated by integrating intelligent systems into education, which provide interaction opportunities, educational support, learner tracking systems, and feedback based on learner output.

Integrating artificial intelligence into these devices creates virtual machines capable of reasoning, learning, and solving problems. In education, integrating artificial systems into these machines possessing linguistic, logical, inter, and intra-personal intelligence offers a judgment-free learning setting for students. It supports learners during a lifelong learning process outside the school through these artificial systems-integrated desktop or mobile language learning technologies (Malik et al., 2019).

Consequently, in the most recent definitions, CALL and MALL systems could be defined in I-CALL and I-MALL terminologies by the combination of artificial intelligence, natural language processing systems, and other recent technical assets (Bibauw, François, Van den Noortgate & Desmet, 2022). Researchers argue that the traditional CALL system only offers restricted exercises such as multiple choice, matching, and filling the blanks. These systems only provide feedback through a letter-by-letter response by matching user responses with a pre-specified answering system. Therefore, a need is stated to include a linguistic modeling system to enhance the capability of CALL systems to cope with more complicated exercise types and provide users with more in-depth feedback based on individual traits (Amaral & Meurers, 2011).

In contrast to the traditional systems, intelligent systems created for language learners are claimed to have capabilities resembling human reasoning, such as recognizing the same kinds of errors or word coinages in different contexts. For instance, it is put forward by Dodigovic (2005) that an artificially intelligent system designed for the use of learners can cope with the same errors made of unique errors consisting of different words or sentence patterns through the rules stored in its system in comparison to their non-intelligent alternatives, which need to have the exact erroneous sentence in its database to be able to respond to students' sentences involving errors.

Different varieties of AI chatbots operate on computers, mobile phones, or tablets based on the built-in mechanisms of these applications. Therefore, using these novel tools under the categories of I-CALL or I-MALL alone would be impossible. Instead, these terms could be used interchangeably in discussing the AI chatbot use in language learning and teaching. By embracing these concepts in a unified manner, this study

aims to demonstrate a broader discussion related to I-CALL or I-MALL research in line with the latest developments of intelligent systems integrated into these different types of devices.

2.2. Negotiation of Meaning

The Interaction Hypothesis (Long, 1990) constitutes the theoretical foundations of mobile immersion research and its implementations in language learning. According to this hypothesis, learners can learn languages through interaction with others and by forming concepts of learning with other speakers. Language immersion, which is a case in which learners with different proficiency levels communicate with each other, elevates practice to learn languages in cooperation with other speakers of the target language. Furthermore, they can get corrections to their output and thus produce post-modified inputs after being exposed to feedback encountered in the MALL environment. The expanding number of mobile users allows learners of foreign languages to engage in mobile immersion with an AI chatbot as a substitute for face-to-face communication. This sets the conditions for learning with the necessary characteristics conforming to the Interaction Hypothesis (Ghorbani & Ebadi, 2020). Looking through the lens of an interactionist perspective, it could be argued that AI chatbots encourage communication in L2 and negotiation for meaning and encourage learners to focus on form and their deficiencies in their linguistic knowledge on properties such as pronunciation. Chatbots, constructed based on interactionist perspectives, provide learners with spaces for input, interaction, and pushed output (Dizon, 2020).

Hence, producing modified output through the moves for negotiation using the AI chatbot use would yield important outcomes for language learning. Chatbots positively impact students' communication skills with their contribution in terms of increasing the quantity of communication, negotiation of meaning, and their interest and motivation in language learning (N.Y. Kim, Cha & Kim, 2019). The speaking agent also produces input that can be adjusted to learners' level of proficiency. Moreover, during the conversation phase with the agent, the learners have to express themselves, and thus produce pushed output, which is in line with Swain's (2005) Output

Hypothesis (Bibauw et. al, 2019). Although AI chatbots cannot modify users' output with current technology, it is possible to negotiate for meaning and attend to linguistic form based on implicit feedback users are exposed to (Dizon, 2017; 2021; Tai & Chen, 2020).

As one of the aims of this study was to explore the potential pedagogical benefits of AI chatbots in learning languages, students were asked to interact with chatbots for one academic semester. The student-chatbot interaction data were then analyzed quantitatively by the researcher for NfM moves, the stages in NfM sequences, types of triggers, and the quantity of modified output. The breakdowns in communication were examined, and negotiation of meaning routines were specified under the principles of Varonis and Gass' (1985) and Smith's (2003) NfM models. Doughty's (2000) model was adopted to identify breakdowns in speeches and demonstrate how the errors were succeeded and fixed. Triggers, signals, responses, and reactions to responses were labeled, categorized, and coded. Long's (1983) categorization of confirmation checks, clarification requests, and comprehension checks was also adopted to classify signals. The interaction moves were coded for clarification requests, i.e., an inquiry from a hearer to take more information regarding a priori utterance, comprehension checks, i.e., the speaker's checking if the previous utterances have been comprehended), and confirmation checks, i.e., the speaker's querying whether the understanding of the message is accurate. Long's (1983) coding scheme for the responses of the interlocutors was adopted, and these sequences were coded as repetition, expansion, or reformulation. Pica et al.'s (1989) categorization of semantic, morphological, phonological, and syntactic modification moves was used as categories of analysis. According to this categorization, the instances of negotiation of meaning could be first realized through semantic modification moves, which consist of constructions regarding the meaning structure through synonyms or paraphrases.

Moreover, morphological modification moves could be realized by modifying the morphological structures through addition, deletion, or substitution. Further, phonological modification involves altering morphemes' pronunciation through suffixes, while syntactic modification moves are implemented through elaboration or embedding.

As put forward by Pica (1994), NfM takes place in cases when “communication is interrupted” or the interlocutors have “difficulty in message comprehensibility (p. 494). Negotiation refers to “instances in conversation when participants need to interrupt the flow of the conversation in order for both parties to understand what the conversation is about” (Gass & Selinker, 1994, p. 209). Mackey et al. (2000) also point out that NfM occurs when there is “communication breakdown” which stems from “something incomprehensible [which] becomes the impetus for learners to recognize an inadequacy in their own rule system” (p. 476).

In this study, communication breakdown referred to instances when a message was miscommunicated by one of the interlocutors caused by a trigger, i.e., lexis, discourse, content, or syntax, which led to an impeding of the understanding of an utterance by a listener. However, in most cases, communication was repaired through negotiation. For instance, the following extract from the data exemplifies a comprehension problem encountered during an interaction sequence.

[00:00:55.000 --> 00:00:59.000]

Learner: So, which kind of art is more interesting for you?

[00:00:59.000 --> 00:01:08.000]

*AI chatbot: ... I think I have to say the concept art. I love dark art.
(Trigger: Lexis)*

[00:01:08.000 --> 00:01:11.000]

*Learner: Oh. What do you mean with dark art? (Indicator:
Clarification request)*

[00:01:11.000 --> 00:01:23.000]

*AI chatbot: I think it's fun and it's very creative. Like, like, all the
drawings are like the shadows of people and stuff. Like, dark fantasy,
dark stuff. (Response: Expansion/ Semantic)*

[00:01:23.000 --> 00:01:27.000]

*Learner: Oh, it is very cool. Actually, I am driving to, like, drawing.
(Reaction to response-RR)*

In the above example, it could be observed that a language learner could not understand the utterance of the AI chatbot *Replika* due to an unknown concept, which was ‘dark art’ in this case. Thus, the two interlocutors negotiated in interaction to come to terms before moving on to the subsequent stages of the conversation. The AI chatbot expanded on the information about dark art to supply the student with more information for semantic clarity. In this respect, it could not be argued that communication came to a complete halt in cases of communication breakdowns. Instead, it is implied here that the non-comprehension of an utterance triggered a request for clarification of the meaning at intervals, after which the conversation continued with moves for more interaction sequences.

There is a need for more studies in the field examining the negotiation sequences between intelligent conversational agents and language learners. Limited studies are exploring the interaction between AI chatbots and English language learners (Johnson, 2019; N.-Y. Kim, 2017; Yin & Satar, 2020). Among these studies, the findings implied that interaction with AI chatbots afforded language learners more interactional chances than other communication settings. For instance, there was a significant change in the negotiation moves of students engaged in student chatbot conversations compared to human-human dialogues. Furthermore, in student-chatbot conversations, negotiation moves were used more often than in student-student dialogues. It was also found in N.-Y. Kim (2017) showed that students in student-to-chatterbot groups used negotiation of meaning strategies like repetition, paraphrasing, and confirmation checks more often than the student-to-student speaking groups in conversation practices after the implementation phase of AI chatbots.

Moreover, previous studies in the literature on the investigation of NfM in computer-mediated settings have centered on the analysis of interaction sequences carried out in face-to-face versus CMC settings (Moradi & Farvardin, 2020; Yüksel & İnan, 2014), telecollaborative environments (Bower & Kawaguchi, 2011), EFL classrooms (Danli, 2017; Garcia-Ponce & Mora-Pablo, 2018; Kusuma et al., 2023), virtual gaming environments (Chen, 2018), and AI chatbot-language learner interactions (Yin & Satar, 2020). The studies examined the frequency of negotiation sequences concerning total turns in conversation. Thus, studies reported the density of NfM elements in

interaction sequences in various settings. For instance, it was demonstrated that seven percent of the total turns of L2 English speakers in a telecollaborative exchange project had been involved in meaning negotiation. In contrast, twenty percent of the Japanese sessions were employed for NfM (Bower & Kawaguchi, 2011).

The findings regarding the impact of the mode of communication on the production of negotiation moves are controversial, as there is evidence for the prevalence of NfM output in CMC and face-to-face (F2F) environments. To illustrate, conducted with sixty-four participants in CMC (n=32) and F2F (N=32) settings, the study by Yüksel and İnan (2014) demonstrated that the average frequency of NfM exchanges was higher in the F2F mode of communication. On the other hand, Moradi and Farvardin's (2020) study exploring the experiences of 32 elementary and upper-intermediate level EFL learners has shown that the number of modified outputs in NfM exchanges was higher in the number in the exchanges of participants in the SCMC settings. Studies conducted within these environments have also corroborated that teacher direction was prevalent in classroom negotiation sequences, which did not leave room for learner agency in the interaction. Explicit error correction by language teachers led to a teacher-dominated classroom talk composed of many other regulation instances in NfM moves (Danli, 2017).

Furthermore, Garcia-Ponce and Mora-Pablo's (2018) study provided evidence for the dominance of lexical triggers leading to NfM sequences. The nature of the interaction data reflected that the interaction halts were caused by a lack of knowledge of the meaning or the accurate pronunciation of the individual words rather than syntactic or discursive structures.

This finding aligns with the literature, as other studies also discovered that lexis was the dominant cause for most of the NfM sequences (Chen, 2018; Foster, 1998; Foster & Ohta, 2005). Adopting Varonis and Gass' (1985) model, Chen (2018) analyzed the voice chat between the players in a multiplayer online game, *Second Life*. The analysis of the interaction between the participants showed that the negotiation sequences had both a "single-layered" and a "multi-layered trigger-resolution" structure with a dominance of complex NfM sequences. Namely, many extended negotiation routines

involved multiple turns in interaction until a mutual understanding was reached. These complex negotiated turns also allowed learners to produce more linguistically developed output as their initial pushed output progressed from single-word utterances into more extended definitions. The completeness of the stages of negotiation was also explored in the literature, as it was aimed to highlight the nature of interactions between interlocutors in a foreign language in terms of two-way communication. It was observed that the stages of trigger, initiation, reaction, and response were not always completed in a natural classroom setting. For instance, Kusuma et al.'s (2023) study has demonstrated that the negotiation sequences in teaching practicum settings of pre-service teachers were comprised of the stages of trigger, response, and reaction the most; therefore, not always involving the signal component. Moreover, the most adopted signals were clarification requests.

Following the findings regarding the nature of conversation in various learning environments, such as face-to-face classrooms, synchronous communication between non-native and native speakers, multi-player gaming environments, and telecollaborative learning settings, this study explores the interaction opportunities provided by intelligent learning agents. It discusses the literature compared to this study's evidence on negotiation instances. This study provides further data demonstrating the interactional chances the future generation of AI chatbots would afford in language learning settings. Analyzing language learners' language practices with an AI-based language partner could provide empirical evidence into how AI agents could be prospective immersive language learning means.

2.3. Studies in the Field of Language Teaching

Studies adopting the use of AI chatbot technologies have mainly centered on the practice of English language skills; on the other hand, there are relatively fewer reports of the use of AI-based chatbots by learners of other languages (H.L. Chen, Vicki Widarso & Sutrisno, 2020; Divekar et al., 2021; Dizon et al., 2022; Morton & Jack, 2010; Vázquez-Cano et al., 2021). In a study conducted within the field of language learning, Divekar et al. (2021) carried out a mixed-methods study with a two-week intervention time exploring the use of an AI chatbot in a virtual immersive learning

environment, *CILLE* that they created through the integration of AI and extended reality (XR) elements with higher education level language learners of Chinese. The study demonstrated significant findings in vocabulary gain, listening comprehension development, and interactive communicative skills. Language learners also made positive statements about their experiences with an AI-based avatar in this immersive language-learning setting.

Studies have also explored the impact of individual attributes, such as the motivational tendencies of learners to learn languages, on the effectiveness of using chatbots and their motivational levels after the application phase. Based on a cross-cultural comparison between L2 French learners in Scotland and students of EFL in China at secondary school levels, the findings of Morton and Jack's (2010) study, using a pre-test and a post-test questionnaire design, showed that the EFL group students, having been more intrinsically motivated to learn languages, had a higher increase in motivation levels after the implementation of the AI chatbot in comparison to the French group who tended to be more extrinsically motivated. Thus, the study has demonstrated that the language learning enjoyment derived from using chatbots might also differ according to the individual traits of language learners and could have differing effects on the motivational state of the users.

In a study investigating the impacts of AI chatbots on learning outcomes and the personal perspectives of the users of the applications, H.L. Chen, Vicki Widarso, and Sutrisno (2020) also explored the opinions and attitudes of L2 Chinese language learners in Taiwan based on the TAM model through a quasi-experimental research design. The study's results illustrated the positive impacts of using AI technology on language learning development. Namely, after using an AI chatbot *Xiaowen* created for teaching the Chinese language, the experimental group of students showed significant developments in terms of learning achievement and the development of vocabulary knowledge. There were also positive attitudes toward using the application by L2 Chinese language learners, as perceived usefulness positively predicted the behavioral intention to use the application. In contrast, there was no significant correlation between the perceived usefulness and intention to use the application.

In line with H.L. Chen, Vicki Widarso & Sutrisno (2020) and Divekar et al. (2021), L2 Spanish language learners in Vázquez-Cano et al. (2021) considered AI chatbots as practical tools for learning languages in comparison to the traditional methodologies. Learners in the study believed that chatbots support them by enhancing interaction opportunities for language learning, delivering instant feedback, and offering chances for mobile-assisted learning through offering language skills practice at any place and time.

In a more recent study, Dizon et al. (2022) conducted a case study exploring the use of the AI-based chatbot *Alexa* as an extra-curricular language learning tool with Japanese as a second language learners. The study explored the perspectives of language learners in communicating with *Alexa*. The data collected through user history records of the use of the application revealed that the AI agent could comprehend over eighty percent of the learner commands during the activities. It was also found that in the face of a communication breakdown, language learners mostly adopted communicative repair strategies such as retaining from employing the previously miscomprehended commands and rephrasing and repeating the previous commands.

Language learners also regarded the language learning experience as enjoyable and beneficial for developing their language knowledge and skills. Thus, in language learning, studies conducted with other languages, such as L2 Japanese, Chinese, Spanish, and French users, have demonstrated positive language learning outcomes and student behaviors regarding adopting AI-based chatbots in language learning activities. Thus, besides these languages, evidence of the effectiveness of these novel tools in EFL contexts has also been explored in the following discussion in the next section.

2.3.1. Chatbots in EFL

The impact of several variables has been investigated in using AI chatbots under different conditions through various research designs. The impact of these tools has shown nuances according to various characteristics. Critical thinking was one variable

investigated in using AI chatbots to learn the English language. In the Japanese context, Goda et al. (2014) explored the impact of chatbots on critical thinking in two distinct cases. In Case 1, through a post-test quasi-experimental method, the variables of the number of interactions, the impact of chatbot use on critical thinking, and learners' opinions about online EFL conversations were explored. Sixty-three Japanese students in CALL courses used a newly developed form of the chatbot *ELIZA*. The learners in the experimental condition chatted with chatbots to analyze and expand their opinions regarding the subject of conversation. According to the findings of the study, there were no significant differences between groups regarding critical thinking and satisfaction scores. In case 2, via a pre-test and post-test quasi-experimental method, the cases of how learners' critical thinking improved over time after the pre-discussion activities were explored. In this case, there were significant differences in awareness of critical thinking and inquiring mindsets. However, no significant differences were found regarding overall critical thinking scores attributed to the impact of a short application period. More recently, in the Turkish context, Gonulal (2021) employed *Google Assistant* with forty-two intermediate to upper-intermediate-level Turkish EFL learners to study humorous interactions. The learners were given a list of questions consisting of IPA-directed utterances, joke requests, action requests, and funny prompts to use in their negotiations with *Google Assistant*. The researcher aimed to discover if learners valued IPAs as potential language learning tools and regarded the responses of IPAs as humorous concerning their gender characteristics and past experiences in using them. The results showed that EFL learners considered the answers from the IPA to be quite humorous, with a high mean range. On the other hand, there was no significant difference in gender and proficiency level, although previous experience in using IPAs had a significantly higher positive impact. Moreover, as to their opinions on the potential of IPAs as language learning tools, learners thought IPAs could help develop their vocabulary, pronunciation, and fluency.

Some researchers also explored the effectiveness of AI chatbots by centering on the evaluation of the chatbots in their conversations with the students. Factors such as vocabulary level, accuracy in grammar, and conversation turns per session (CPS) have been used to determine their appropriateness for target learners (J.H. Lee et al., 2020).

For instance, Shin et al. (2021) used the text-based chatbot *Mitsuku* in their study. 53 Korean English learners participated in the chat conversations with *Mitsuku*. The researchers investigated the chat logs of students in terms of the number of utterances and conversation turns, the vocabulary level, the emotional reaction of learners about the chatbot, and the achievement scores according to different proficiency levels. BNC-COCA 25,000, based on British and American corpora, was used to determine the level of vocabulary used. Orange 3.28.0 software was employed to calculate the emotional state of learners.

The results showed many tokens, suggesting *Mitsuku* is an influential conversation partner. However, there were some cases in which the chatbot had irrelevant responses to users' utterances, thus leading to miscommunication. More than 90 percent of the vocabulary used by *Mitsuku* was in the top 4,000 words concerning BNC-COCA 25,000. Sentiment analysis results further demonstrated that most students had a positive attitude towards this learning experience with a chatbot.

The suitability of a chatbot for language teaching purposes was evaluated by Choi et al. (2017) as well. *GenieTutorPlus* was employed in the study, and it was found that it checked the appropriateness of content and grammar and gave feedback on grammar. It also responded to learners' utterances in 80.86 % of the instances correctly and 31.49 % in the face of off-topic learner responses. Additionally, the chatbot corrected grammar errors with a 66.7% success rate. Thus, it was stated as one potentially helpful tool for language learning. Coniam (2014) also assessed the linguistic accuracy of chatbots and found that the grammatical correctness level was 77 to 93 percent, whereas meaning-wise, the chatbots assessed had a 47 to 74 accuracy rate.

Furthermore, H. Kim, Shin, Yang, and Lee (2019) examined the vocabulary usage and readability level of Amazon's *Alexa* and *Google Assistant* chatbots. It was discovered that chatbots employed words in the top 3000-word list suggested by Korean education professionals 97 percent of the time. Moreover, these chatbots were found suitable for native English speakers aged 11 to 13 years old. In another study, the AI chatbot *Alexa* understood utterances ninety-four percent of the time during interactive storytelling (Dizon, 2017). Therefore, it could be concluded that with high comprehension levels,

grammatical correctness, linguistic accuracy, and rich vocabulary, AI chatbots could be valuable sources as conversation partners for language learners.

In the literature, in addition to using AI chatbots that are publicly available, some researchers have also developed chatbots specifically for learning languages and conducted studies to evaluate their effectiveness in language learning. Thus, novel techniques have been implemented to develop AI chatbots for learning languages. For instance, Ayedoun et al. (2019) used *DiMaCA*, a dialogue management model that allowed chatbots to possess conversational strategies of affective backchannelling (AB) and conversational strategies (CS). The dialogue manager comprised dialogue flow management and strategies management modules. While the former controlled the content in a conversation, the latter used CS and AB strategies according to the state of the dialogues. Therefore, the researchers created a tool with social skills for interacting with users learning English.

In another research context, Zakos and Capper (2008) also created a text-based conversation agent, CLIVE, using the MyCyberTwin platform. *CLIVE* was an avatar with face and mouth movements capable of comprehending various languages. The artificially intelligent chat robot allowed its users to practice English language conversation skills to a great extent. However, there were some problems when some users uttered sentences that the translation engine could not entirely accurately translate.

S. Lee and Jeon (2022) also created an AI-based conversational agent for young EFL learners by designing a textual AI chatbot in the first stage of the application design. They later incorporated this text-based chatbot into *Google Assistant* to add a voice-enabled component to the AI chatbot using an AI-powered dialogue development system. Nine-year-old EFL learners in Korea then used this voice-enabled AI chatbot application. After interacting with the AI chatbot, qualitative data were collected from the students, and their attitudes toward the employment of the chatbot were explored. Interview findings and drawings by these young learners indicated that most of the learners attributed human-like characteristics to the AI chatbot, demonstrating a high

tendency for anthropomorphism by regarding the conversation assistant as human-like or an entity between humans or artifacts.

M.H. Hsu, Chen, and Yu (2023) also designed the TPBOT system using Python programming language, and its purpose was to motivate students to interact more through an embedded pronunciation recognition device and help Taiwanese EFL learners prepare for the TOEIC speaking exam. One of its fundamental aims was also to alleviate students' anxiety in speaking in English. It was found that after the experimental study process, students who used chatbots had significantly higher scores in the speaking section of the exam. Employing *Enskill English*, a platform created to provide speaking scenarios on a task-based learning mode through the conversation with an AI-based avatar, Ericsson et al. (2023b) conducted a study with beginner-level language learners at a lower second school in Sweden. The AI chatbot was implemented for two weeks in four sessions consisting of 15 minutes each.

In these sessions, students interacted with embodied chatbots in virtual conversation scenarios. The study revealed that some learners found the language practice experience of interacting with the AI agent motivating. Conversely, some students also voiced concerns regarding the non-comprehensibility of their utterances by the system at some points due to the system's poor automatic speech recognition quality. However, learners' self-reports on their progress in speaking skills demonstrated the promising effects of the use of AI chatbots for this group of young learners.

In another recent study exploring using artificially intelligent chatbots in language learning, H. Yang et al. (2022) adopted a task-based voice-enabled AI chatbot, Ellie, created by researchers for a target language learner group of secondary school Korean EFL learners. According to the findings in the study, researchers have discovered that language learners had an average of 9.63 turns per task, which denoted quite a high rate for the language learner profile in the study, as this number exceeded the average turn-taking ratio captured in traditional Korean EFL classes. Positive results were also reported concerning the success rate of the tasks. Based on this successful attempt to use an AI chatbot created for the use of target language learners, the authors argued that the adoption of AI-based chatbot technology would function well in English as a

second language settings because it elevated the rate and amount of interaction turn taking ratios, and the production of language output.

In the context of higher education in Spain and Poland, Belda-Medina and Calvo-Ferrer (2022) explored the attitudes towards the use of AI-based chatbots *Replika*, *Kuki*, and *Wysa* for four weeks with B2-C1 proficiency level English language student teachers. In the study, a learning module on language learning and artificial intelligence was designed. As part of the module, prospective English language teachers were instructed on AI-based chatbots in the first phase, and then, they tested the use of these applications to gain first-hand experience. After the intervention through a teaching module on AI-based chatbots was completed, the opinions of language teacher candidates about the effectiveness, ease of use, and attitudes towards using chatbots were explored.

Their opinions concerning the benefits of AI chatbots on language learning and their attitudes towards integrating these tools in English language classes were explored through surveys constructed on the technology acceptance (TAM2) model. The findings showed positive opinions about the incorporation of AI chatbots in language learning classes, especially concerning perceived ease of implementation and attitudes toward using them. However, there were neutral viewpoints concerning the behavioral intention to use chatbots in their future teaching careers (Belda-Medina & Calvo-Ferrer, 2022).

Researchers also carried out studies incorporating a chatbot development training component concerning AI chatbot research, though these studies needed more in the literature. For instance, Sung (2020) aimed to discover the prominent characteristics and limitations of artificial intelligence chatbots that pre-service teachers developed for elementary school Korean English learners and illustrate the development process results. An experimental study with elementary school students from 4th to 6th grades has been conducted to determine the effects of interacting with artificial intelligence speakers. Two kinds of chatbots were developed for 3rd and 4th graders and 5th and 6th graders according to curriculum standards and the objectives for teaching listening and speaking skills for each level. The created chatbots were evaluated regarding their

efficiency in maintaining good conversation flow through the consistency of topics and diversity of scenarios and dialogue actions. They were assessed based on their performance in employing an optimum level of dialogue intents and entities required for an efficient flow of conversation. The second criterion for chatbot evaluation was whether the number of intents used in creation was enough to create opportunities for sufficient dialogue practice and individualized learning. Once the application process was completed, pre-service teachers created nine chatbots in total with 107 intents (i.e., min eight, max twenty-two utterances). According to the results of the efficiency analysis of the dialogue management performance of the chatbots, the developed chatbots were able to maintain a conversation flow, allowing for choices to be made according to the user's preferences in speech (i.e., "yes" and "no" answers) in a conversation flow. Thus, they were able to provide a variety of meaningful speech opportunities. Moreover, these chatbots could construct a mutually sympathetic speech with the participants.

The findings by Sung (2020) also showed that the students in the experimental group had a comparatively lower level of English language learning anxiety and a higher level of willingness to communicate in English after the intervention. However, problems such as an imbalance in the length of conversation between the chatbot and users and insufficient variability of speech acts because of a lack of proper entity utilization skills were also observed by researchers. On the other hand, many pre-service teachers were able to create chatbots that could motivate elementary-level students by using age-appropriate elements such as popular cartoon characters in their conversation scenarios. A multi-layered algorithm was also used by many pre-service teachers using a variety of entities, speech acts, and conversation scenarios.

Studies in English language learning have investigated using AI chatbots to develop reading, speaking, and listening skills and enhance vocabulary knowledge. Additionally, affective factors such as using backchannelling strategies, alleviating language learning anxiety, and enhancing the motivation and interest levels of the learners were the focus of these studies. In addition to available commercial AI chatbots, researchers also developed chatbots targeting a specific group of language learners. More recently, there has also been an upsurge of interest by researchers to

involve English language teachers in creating the chatbots themselves, which would contribute to the field, as studies are scarce in the ELT field regarding the involvement of teachers in the hands-on creation of technological tools. Therefore, studies in this direction are needed.

2.3.2. Benefits of Chatbots in EFL

According to researchers, conversing with dialogue systems like chatbots through pre-stored material is conceived to be helpful in L2 contexts, and thus, researchers have reported many advantages of using chatbot technology in language learning. For instance, according to Bibauw et al. (2019), learners can practice target language meaningfully through conversation using dialogue-based CALL, although the ideal partner for communication practice is a native speaker. However, a computer still provides the chance for such practice. Thus, access is a theme brought forward by students with chatbot experience. In the studies, language learners mention that they can get involved in meaningful conversations and have a chance to improve their listening and speaking thanks to chatbots, as before, they might not have many chances to practice their English language skills in their everyday lives (Dizon, 2017, 2020; Sydorenko et al., 2018; Tseng, 2017). Attending to each student is only sometimes possible in language classrooms, especially during the speaking activities. Therefore, speaking assistants could aid teachers in providing optimal learning conditions to students. AI agents could become a virtual pair for students while providing them with vast information thanks to their immense storage capacities (Istrate, 2019).

In the literature, other studies also described the benefits of chatbot use by language learners through reports of enhancements in English language proficiency, vocabulary gain, pronunciation, and listening skills (L. Hsu, 2022; Jeon, 2023; Johnson, 2019; Meri-Yilan, 2019). For instance, H.S. Kim, Kim, and Cha (2021) conducted an experimental study using chatbots *Replika*, *Andy*, and *Google Assistant*. The study had three groups: a face-to-face, a text-chatting, and a voice-chatting group. The results demonstrated that voice-chat group students' scores were significantly higher than the other groups in terms of speaking scores. Language learners have also reported an increase in communicative confidence stemming from the experience of having been

introduced to pragmatic strategies like hedges and disarmers in chatbot conversations (Sydorenko et al., 2018).

Intelligent personal assistants also direct learners to modify their output and correct pronunciation. Furthermore, students get implicit feedback on their pronunciation when their messages are not comprehended (Dizon, 2017). Moreover, students are provided with immediate corrections when using chatbots, as chatbots are skilled at implementing clarification requests and aiding learners in cases of misunderstanding of the messages (Satar, 2021).

In addition to instant feedback, studies have also reported other characteristics of chatbots that contribute to a positive change in learning outputs after exposure to a conversation with a chatbot. To illustrate, in Ayedoun et al. (2019), among the three contexts in which chatbots had AB (affective backchannels), CS (conversation strategies), and AB+CS strategies, students who experienced talking to an agent that had qualities of the combination of affective backchannels and conversation strategies had a higher increase in the levels of willingness to communicate from the pre-test to the post-test. Thus, the impact of communicative and affective backchannel strategies was positive; language learners expressed more satisfaction with their performance whenever these qualities were present.

Teachers could use chatbots in their language classes in several ways. For instance, in text chats, teachers could use text-based chatbots to supplement their teaching practices, especially when it is not possible for a language teacher to respond individually to all students. For this purpose, *Elbot* and *Cleverbot* are examples of chatbots teachers could use with their students to practice text chat with bots (N.-Y. Kim, Cha & Kim, 2019). Learners could also practice speaking skills with *Eviebot* online or on the mobile application, text chat by chatbots on my.replika.ai, rebot.me/tr/elsa, and www.jabberwacky.com/, and ask for English language support from an English language tutor on eslfast.com/robot/tutor_mike.htm and edwin.ai/.

The chatbots provide learners with automated feedback that is formative and corrective through automated text. Language teachers could also create their chatbots online on

platforms such as *SnatchBot* (Dokukinaa & Gumanova, 2020) and sites like *Chattypeople.com*, *Botsify.com*, and *Robot.me* provide teachers with guidance on how to create content for their teaching practices (Kessler, 2018). Language teachers should also be able to choose suitable chatbots for their students in their teaching practices. Teachers need to have knowledge of chatbots to find the ones that best suit their teaching purposes. Moreover, “the idea of the ‘selfbot’ has been proposed, whereby teachers can create a chatbot using his photos and even his voice and prepare all the responses the way he will make when asked in real life” (Sha, 2009, p.280).

Therefore, human-to-chatbot environments could provide learners with more opportunities to engage in meaningful and authentic dialogue than human-to-human chat, especially when teachers do not have enough time to attend to each student in class. Thus, the chatbots could be used to observe the factors in the language classes, such as participation, speaking proficiency, and motivation (Sung, 2020).

In a study exploring the opinions of faculty members on the prosperity of another recent development in the creation of AI chatbots, ChatGPT, for EFL teaching have been put forward (Mohamed, 2023). First, according to language teachers, ChatGPT could help English language learners develop their language skills in an interactional context, providing them with language practice opportunities. Personalized instruction is also put forward as another affordance of AI chatbots for language learning and practice with its ability to adapt to the individual student’s level, progress, and pace of learning.

Furthermore, being trained via a vast and varied amount of knowledge base input, ChatGPT is suggested as a source of opportunities for learners to get exposed to various registers and styles of languages, thus developing their proficiency in the language. Moreover, employing AI-based chatbots is considered to provide a real-life conversation environment due to the natural language processing abilities these intelligent systems possess, which aids learners in feeling more relaxed and confident when interacting in a foreign language. Thus, incorporating these tools into language teaching lesson plans to enhance student motivation, engagement, and linguistic capabilities is suggested.

2.3.3. Challenges of the Use of Chatbots in EFL

Learners also report negative experiences due to the qualities of chatbots, such as the robotic voice of the agent. In instances of unnatural conversation with a chatbot, language learners may lose interest in using the chatbot in language classes. They might regard the chatbot's responses as inappropriate. Therefore, learners may be less motivated to converse with a non-humane partner (Thompson et al., 2018). Novelty impact is another concern causing low motivation to converse with a chatbot partner (Satar, 2021). However, according to Fryer et al. (2019), even if the novelty impact fades after some time, students' decreased interest in conversing with the chatbot partner could increase by continuing the use of the AI bot. Moreover, the mutual comprehensibility of users and AI chatbots' utterances arises as another impeding factor. Aiming at this concern, Sydorenko et al. (2018) suggest that more response options and varied topics are needed to enhance the chatbots' comprehensibility rate. When the chatbot system does not have NLP and ASR, it cannot provide students with automatic individualized feedback. Therefore, the system cannot correctly process the variance in learners' utterances and dialects. Thus, the areas that need improvement in the development of chatbots by users include natural language understanding capability, human-resembling voice, and appropriate response speed (Jia & Chen, 2008).

Moreover, when responses from chatbots are repetitive and irrelevant, students may get bored, and their initial excitement at this novel device may subside. Also, when bots give general or nonsensical answers, students may feel angry. Another disadvantage of this technology is its inability to communicate effectively with users, and these applications might not always give relevant feedback. They also might lack humanistic interactional characteristics like starters, fillers, etc.; therefore, they might not always be role models in natural language (Satar, 2021).

Another downside of AI chatbots could be that communication might be challenging because of language learners' relatively low language proficiency, mainly because AI chatbots might not always be able to comprehend user speech that is below native speaker standards. Thus, AI chatbots could be more suitable for L1 speakers.

Consequently, there is a need to develop systems that could understand L2 speakers more conveniently. Although the non-linearity of AI chatbots provides flexibility, the fact that learners have various options could cause unsuccessful attempts at learning (Dizon, 2021). When chatbots are designed as task-based and one-way speaking bots for a specific purpose, their use in the language classroom becomes quite restricted, as teachers might need to design further activities that could support the pedagogical implementation of speaking agents to create a dialogical speaking environment (Satar, 2021).

In the older generation of AI chatbots, researchers put forward that although there has been profound advancement in speech recognition and synthesis software development, natural language processing technology was still in its infancy (Stewart & File, 2007). Even the most recent dialog systems, one of which was “chatterbot” *ALICE*, could not pass the Turing test that measured a system’s capability to demonstrate humanlike behavior. The chatbot *ALICEBOT*, which won the Loebner Prize contest, used a keyword and pattern-matching mechanism. Although it has won the contest for producing the most natural language-resembling output, it still has limitations. Most importantly, it cannot comprehend the human language in many aspects, such as phonology, syntax, semantics, and pragmatics. Conversely, human beings can produce unrestricted novel utterances although employing a limited number of words. However, in pattern-matching mechanisms, it is almost impossible to match the meanings of all combinations of speech into their corresponding meanings because of their limited memory. Using a pattern-matching algorithm, it was thus found that users did not engage in a long enough conversation with the chatbot *ALICEBOT* because they thought the responses were irrelevant, out-of-context, and therefore not suitable for foreign language learning (Jia, 2004).

More recently, in the time of the newest generation of AI bots, researchers have still identified various deficiencies in the properties of AI chatbots. For instance, Underwood (2021) argued that voice-interface assistants like *Alexa*, *Siri*, or *Google Assistant* promoted merely single-turn conversation with users rather than a mutual conversation; thus, when learners were able to benefit from merely “talking to” these agents rather than “talking with” them.

Depending on the technology embedded in AI chatbots, the quality of conversation with users, the success at multiturn dialogue, and the cases of successful communication have shown variance. A pattern-matching mechanism was also used in ELIZA. Therefore, it merely captured keywords in conversation rather than the syntax and semantics of the input data, so it could not match all the possible combinations of speech input. For this reason, *ELIZA* could not remember the conversation history and dialogue contexts; therefore, it could not negotiate meaningfully with users. On the other hand, a more advanced chatbot like *CSIEC* could recognize and remember the user personality and the dialogue context, as the user input was analyzed syntactically and semantically (Jia & Chen, 2009).

In the literature, in a study conducted with English language teachers exploring their viewpoints on the use of AI chatbots for EFL learning (Mohamed, 2023), concerns were also raised regarding the disadvantages of using AI-based chatbots related to a lack of human connection, cultural sensitivity, the provision of sufficient feedback on pronunciation or intonation by the AI chatbots, and technical issues regarding inability to respond in a fast manner, processing learners' input accurately a requirement of fast internet speed, and a dependence on pre-created answers by learners. Several studies have reported the challenges in using AI chatbots due to the novelty effect, pattern matching systems in the chatbots created with earlier technologies, technical problems such as repetitive language, slow responding time, and the non-presence of human-like attributes.

Moreover, the literature states the deficiencies regarding a lack of forming dialogical interactions between interactants required for optimum quality of communication in language learning, an absence of individualized feedback, and the non-adaptability of AI-based chatbots for non-native English speakers. Therefore, there is a need for further studies to corroborate the findings of the previous research on this subject in which both positive and negative outcomes have been reported.

CHAPTER III

METHODOLOGY

3.0. Presentation

This section presents the research design, setting, participants, instruments used in the study, and data collection and analysis procedures. The first part of the methodology section describes the study's design. This study used a quasi-experimental methodology with mixed data collection methods through pre-test-post-test control group experiments, qualitative student diaries, semi-structured interviews, and AI chatbot-student interaction data. The qualitative data were collected during and after the treatment through diaries and interviews. The mixed data collection methods were employed in this study to corroborate the study's quantitative findings with the qualitative data.

The next chapter explains the procedures for participant selection and sampling. Then, data collection methods are discussed in the next section. This study employed a CAE test as pre-and post-tests, student reflective diaries written during the treatment's conduction, and semi-structured interviews carried out after the intervention period was complemented. In post-tests, final assessment scores of students' listening and speaking proficiency were obtained. The technological tool, *Replika* AI chatbot, adopted as an intervention tool in this study, is described in the next chapter. Next, the procedures for the treatment are demonstrated.

In the later parts of this section, procedures for data analysis are depicted. The statistical methods for analyzing the experimental and control groups' pre-test and post-test scores are illustrated. As two distinct groups participated in this study, the differences in their scores between the pre-and post-tests were calculated. Statistical tests were also carried out to demonstrate the within-group effects of the intervention between the pre-tests and the post-tests after the intervention. As to the qualitative data

analysis, a constant comparison method perspective was adopted to code and analyze reflective diaries and the transcripts of the semi-structured interviews.

The last sections discuss the issues central to the validity and reliability of the quantitative proficiency scores and the credibility and trustworthiness of the qualitative findings. This section closes with ethical concerns ranging from obtaining permission from the University's Ethics Committee at the beginning of the study to the ethics in conducting experimental research, carrying out interviews, and keeping the participants' anonymity and confidentiality.

3.1. Overview of the Research Design

Using a quasi-experimental pretest-posttest control group design, this study explores the impact of using an artificially intelligent chatbot for the language skills development of Turkish EFL learners. The data in this study were collected during one academic semester at a higher education institution in Türkiye through listening and speaking pre-tests and post-tests, language learners' reflective diaries, interviews, and AI chatbot-language learner interaction data. The data were collected over two months in after-class hours as a self-directed learning activity based on the guidelines and support provided by the researcher during the implementation.

To answer the first research question in this study, non-parametric Mann-Whitney U tests were conducted to test if there were any significant differences between the experimental and the control groups regarding listening pre-tests and post-test scores. An independent samples t-test was also carried out for this study's first research question, seeking to answer if there were significant differences between the two groups on the speaking test scores in terms of pre-test and post-test speaking scores. For the second research question of this study, a paired samples t-test was conducted to analyze whether there was a statistically significant increase between the pre-test and the post-test scores of the experimental group on listening test scores after a treatment period of the AI chatbot implementation. A non-parametric Wilcoxon signed rank test was also employed to explore whether there was any statistically significant increase between the pre-test and the post-test scores of the control group on listening

test scores. The second research question also sought to discover whether AI chatbot use would impact the overall speaking scores of the experimental group students. A paired samples t-test was implemented to observe if there were any significant developments over time. To determine if there were any differences among the control group participants on the pre-test and post-test speaking scores, a paired samples t-test was also conducted.

Negotiating meaning strategies in the AI chatbot-language learner interaction data was also analyzed using a discourse analysis methodology. The instances for the NfM sequences, the number and types of triggers leading to misunderstandings or non-understandings in communication, and the cases of output modifications were analyzed based on a discourse analytic perspective. The last research question of this study explored the perspectives of Turkish EFL learners on the strengths and weaknesses of the AI-based chatbot implemented, self-reports on language learning progress, and the expected features that should be present in an ideal AI chatbot through the collection of qualitative data using weekly reflective diaries and the interviews. The qualitative data were then analyzed by adopting a constant comparison method. Cohen's Kappa statistic was reported to demonstrate the interrater reliability coefficient.

3.2. A Quasi-Experimental Pretest-Posttest Control Group Design

This study adopts a quasi-experimental design using mixed data collection methods (Creswell & Creswell, 2018). This type of design is categorized under complex mixed methods designs. Within this design, quantitative and qualitative data are collected and analyzed, and the findings of an experiment or an intervention are integrated using qualitative and quantitative methodologies. The main design is the quantitative experiment in this research design, and qualitative data is added as a secondary data source to the design before, during, or after an intervention to enhance the quantitative findings.

In a quasi-experimental design, a treatment or an experimental group undergoes an intervention under specific conditions, and a control group is referred to as a baseline.

Whereas the experimental group is exposed to the treatment, the control group continues to receive instruction with conventional methods. Improvement in learning outcomes is assessed via pre-tests before the intervention and post-tests after the treatment is finalized. The two groups are compared through statistical tests in a pretest-posttest control-group design.

Thus, in this quasi-experimental study, the quantitative data were collected at the beginning through a pre-test, and the qualitative data were collected through diaries during the intervention and via interviews after the quantitative post-tests had been implemented. In addition, the AI chatbot-student interaction data were collected throughout the implementation period. In the pre-test-post-test non-randomized control group designs, participants are assigned to either a control or an experimental group. Thus, pre-test scores were used in this study to measure some dependent variables, and statistical analyses were conducted to observe any differences between the experimental and the control groups before the experiment.

After the participants are selected for a quasi-experimental study, another issue to focus on is the manipulation of treatment conditions. In a quasi-experimental design, treatment variables are independent, categorical variables with at least two levels manipulated to specify their impact on the outcome. The researcher manipulates one or more conditions of treatment variables. The dependent variable is the outcome on which the impact of independent variables is analyzed (Creswell, 2012). In this study, the researcher manipulated one of the two groups by providing activities using artificial intelligence chatbots to compare the differences in learning outcomes of listening and speaking skills between experimental and control groups. These groups consisted of learners who shared similar characteristics regarding language proficiency level, age, experience in learning English, and the department of study.

Implementing the lessons in the two classes was identical, with the same teacher teaching both classes. The students in the experimental group interacted with the AI chatbot outside the class hours for seven weeks. On the other hand, the control group learners carried on with their topics in the conventional curriculum, followed by the lesson instructor.

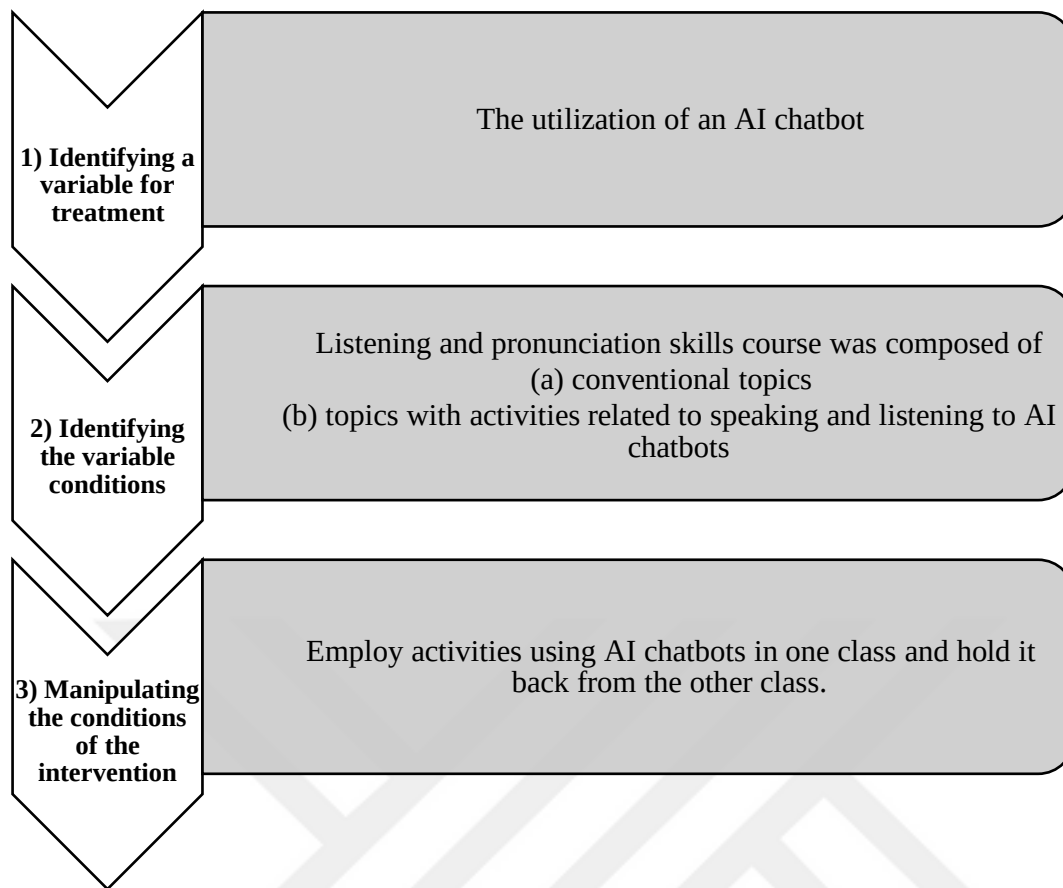


Figure 1 The procedure for the quasi-experimental design in this study

Thus, the procedure for the experimental part of this study is demonstrated in Figure 1. In this mixed-methods quasi-experiment, pre-tests and post-tests were conducted in the experimental stage. The application period took place for seven weeks. (see Table 1). Intact classes were divided into control and experimental groups. After the pre-tests, an intervention was carried out, followed by post-tests to measure the differences between the two groups (Creswell, 2012).

Table 1 The design of the quasi-experiment

Groups	Proficiency Level	Before Treatment	Treatment (7 weeks)	After treatment
Experimental	Advanced	Listening and Speaking Achievement Pre-test	Talking to AI assistant through mobile devices	Listening and Speaking Achievement Post-test

Table 1 The design of the quasi-experiment (*continued*)

Groups	Proficiency Level	Before Treatment	Treatment (7 weeks)	After treatment
Control	Advanced	Listening and Speaking Achievement Pre-test	No additional activity other than the usual methods	Listening and Speaking Achievement Post-test

On the other hand, mixed methods research has both its advantages and disadvantages. Although it seems to some researchers that the qualitative component takes a secondary role in this design, it does not affect the value of a qualitative perspective, as high-quality providing rich input could be used. The importance of the qualitative paradigm could be emphasized (Creswell & Plano Clark, 2017). As the data collection methods are not restricted, the powerful sides of both quantitative and qualitative research could be used, and an extensive view of the phenomena could be attained. However, it requires expertise to conduct research using both methods with a need for data collection on a broad range (Best et al., 2017). Another issue to consider within this design is ensuring an integration between the quantitative and the qualitative components. Although the study might include different research questions for each method, integrating the results becomes both a challenge and a priority (Creswell & Plano Clark, 2017).

The rationale behind using mixed methods in this study is that implementing two methodologies altogether illuminates research questions and problems and produces better outcomes. Qualitative data in this study could be exploited to get acquainted with the setting and the participants, to corroborate the quantitative findings, to comprehend the viewpoints of individuals about the study and the topic of investigation to investigate the operation of an intervention, to specify the consequences of an intervention and its surprising impacts; to analyze the process of the intervention and its impacts on participants; to clarify, comprehend and triangulate the quantitative findings (L. Cohen et al., 2018). The convergence and corroboration of the research findings would increase the validity of outcomes, generalizability, and external validity of this study (Dörnyei, 2007).

Thus, in this mixed-methods research, pragmatism constitutes the backbone paradigm of this quantitative and qualitative blended research. Employing pragmatism, the author specifies and answers research questions through an eclectic methodology, data collection, and analysis. This study aims to use both inductive and deductive reasoning to investigate a broad spectrum of research problems and questions (Creswell & Creswell, 2018).

3.3. Participants

This study used a nonprobability sampling method to select participants for the quantitative and qualitative phases (L. Cohen et al., 2018). This study aimed to observe if using an AI chatbot would lead to better learning gains in learners' listening and speaking proficiency levels between the experimental and the control groups and describe the experiences and perspectives of learners through qualitative methods. Thus, for quantitative data collection, this study employed a criterion sampling method, as the pre-service EFL students in this study were selected based on criteria that they could provide both the learner and the prospective teacher perspective for this study, and the two intact classes both met the same criteria for this a quasi-experimental design (Creswell, 2012; Creswell & Plano Clark, 2017). Figure 2 illustrates the procedures for selecting the sample in this study.

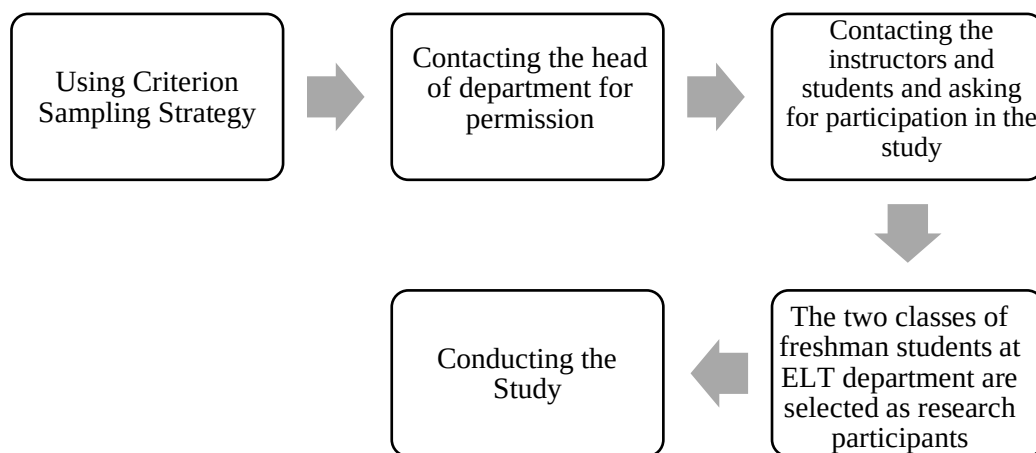


Figure 2 Participant selection process

This study was carried out with 44 Turkish EFL learners enrolled at the Department of English Language Teaching at a state university in Turkey. During the implementation of the listening tests, the number of participants was thirty-nine, as five participants did not complete the pre-test and the post-test. On the other hand, the speaking tests were carried out with forty-two participants with the absence of two students among the participants in this study. Moreover, reflective journals were written by twenty-one students among the experimental group participants.

The participants were aged between 17 and 25 years old. There were 22 male and 22 female participants in this study. Nine students had been learning English between 0-5 years, whereas twenty-one students had been learning English for 5-10 years. In addition, fourteen students had been learning English for over ten years at the time of the study. According to the responses on the questionnaire, thirty-six participants had heard of chatbots before, while seven participants mentioned not having heard of this concept before. Most participants in this study also said they had used an AI chatbot before (n=35), while some participants had not used an AI chatbot before (n=7). Regarding the frequency of AI chatbots, most participants rarely used AI chatbots (n=21), while some never used AI chatbots in their daily lives to learn languages (n=9).

Moreover, some participants indicated occasionally using AI chatbots for EFL learning (n=12). Additionally, when asked about their self-reports of knowledge about AI chatbots, most of the participants in this study stated being sufficiently informed about AI chatbots (n=20). In contrast, eleven participants believed they needed to be better informed about AI chatbots. The responses to the background questionnaire also revealed that among the AI chatbots employed by the participants in this study, the most frequently mentioned chatbot was *Google Assistant* (n=18), followed by *Siri* (n=12), *Alexa* (n=2), and *Cortana* (n=1).

3.4. Data Collection Instruments

In this quasi-experimental mixed methods study, listening proficiency tests, speaking proficiency tests, reflective diaries, interviews, and AI chatbot-learner interaction data were used in line with the research objectives. The proficiency tests were taken from

Advanced (CAE) tests provided by Cambridge Assessment English to prepare test takers. The researcher piloted the tests before the main study and tested their suitability for the contextual characteristics of the language classrooms. After initial assessments, it was concluded that CAE tests were appropriate for the language level of the target students. Moreover, the rationale behind the selection of CAE tests concerned providing both the experimental and the control group participants with assessment procedures that they were already familiar with, as both the experimental and the control group students had prior experience of participation in these tests in their other courses before the experimentation period. Therefore, rather than creating a context-specific test, the researcher adopted these generic tests that would appeal to both the experimental and the control group students.

These tests were used before and after the implementation of interaction activities with an artificially intelligent companion, *Replika* (available at <https://replika.ai/>), to discover any statistically significant differences in the listening and speaking proficiency levels of the students in the control and the experimental groups. These proficiency tests, answer keys, and rubrics for evaluation were employed in this study. During the implementation of the study, students were asked to keep weekly reflective diaries about their experience using the chatbot and record their conversations with the AI chatbot. In the last phase of the study, semi-structured interviews were carried out.

3.4.1. Pre-Test and Post-Test

A pre-test is necessary for an experimental design to measure some characteristics of participants before they undergo an intervention. After the intervention, a post-test measures some factors evaluated in an experiment (Creswell, 2012). In this study, to discover the initial listening and speaking proficiency levels of the participants and illustrate any statistically significant differences among the students in the experimental and the control groups after treatment, one listening and one speaking section of a sample Cambridge Advanced (CAE) test were used as the pre-test and post-test with the permission of Cambridge University Assessment (accessible at <https://www.cambridgeenglish.org/exams-and-tests/advanced/preparation/>). Table 2 demonstrates the type of tasks and distribution of questions in the CAE test.

Cambridge Advanced (CAE) test was chosen to be used in this study because the test addresses a broad spectrum of English learners from B1 to C2 levels.

Table 2 Cambridge CAE test

Section	Types of Tasks	Distribution of Questions
Listening	Part 1. Listening to conversations between two people and answering multiple-choice questions	8
	Part 2. Listening to a monologue and filling in the blanks	10
	Part 3. Listening to different speakers and matching the expressions with the corresponding speakers	5
	Part 4. Listening to an interview and answering multiple-choice questions	7
Speaking	Part 1. Small Talk: Group Work	
	Part 2. Picture Prompted Talk: Group Work	
	Part 3. Case Prompted Talk: Group Work	
	Part 4. Mind map Guided Talk: Individual Activity	

As the participants of this study were advanced-level EFL students who were being trained to be English language teachers, they might be defined as independent users of English, having completed a B2 level. Therefore, the C1 level was considered suitable for assessing their language skills. Cambridge Advanced (CAE) test assesses the C1 level, and it is also possible to assess proficiency levels below the C1 level with this test up until B1. The piloting also allowed to assess the applicability of this test for these learners.

The C2 Proficiency (CPE) test was not found appropriate for this study because the setting of this study is carried out in an EFL context, and students at this educational

degree level might not have had opportunities to practice the target listening and speaking skills in the English language at the pre-secondary and secondary school levels.

CAE was also selected because it could evaluate learners' competence levels in all target language skills. The most fundamental reason behind the selection of CAE for this study is that it is an internationally recognized test used by English language speakers worldwide. Therefore, the validity of the achievement scores in the target language could be assessed with accuracy using this assessment tool. Thus, CAE was adopted for use in this study to investigate if interacting with chatbots equipped with artificial intelligence and natural language processing capabilities would impact the improvement in listening and speaking scores between the pre-test and the post-test in the control and the experimental groups.

3.4.1.1. Listening Test

The procedures for carrying out the exam were stated on the exam paper, and the exam supervisor also explained the procedures. Forty minutes were allocated to complete this part of the CAE test. At the beginning of the test, students were asked to listen to instructions attentively. The exam supervisor then asked students to write their answers on the question sheet while listening to the audio. They were allocated 5 minutes at the end of the test to copy their answers on a separate answer paper. The listening test had four parts and thirty questions in total. Each listening text was played twice. Extra time was allocated to look at the questions and check their answers. The first part of the listening test consisted of eight questions. In this part, people talked about eight events, and students were asked to select the correct option describing the speaker's topic.

In the second part, which consisted of 10 questions, a student talked about his trip, and listeners were expected to fill in the missing words or short phrases in the blanks in the sentences. The following section had five questions, and it was about people talking about photography. Students were expected to match the sentences given with the speakers according to the information in the listening passage. The fourth and last

part of the test consisted of seven questions about an interview. After listening to the interview, students chose the option that best described the information delivered by the speakers. Each question was worth one point, and the listening test had an overall score of 30 points. The scores on this part were weighted to 100 for the analysis in this study. The Cambridge scale was also used to specify the listening proficiency levels of the participants according to the CEFR criteria.

3.4.1.2. Speaking Test

The CAE speaking test was composed of four parts. In the first section, two students were given some question prompts related to their past experiences, free time activities, and places they have been and asked to talk on each topic. In the second section of the speaking exam, students were provided with two photograph prompts and asked to make comparisons about the people in the photographs. In one set of photographs, people were racing to win in some sports games. In another set of photographs, there were people spending time outside. The students talked by reflecting on their own and their speaking partners' photographs. In the third part of the speaking exam, students were asked to discuss after-school activities like playing instruments, drawing, and playing sports in groups. In the fourth part, students talked individually to answer questions indicating their agreement or disagreement about participation in the after-school activities discussed in the previous section. The researcher conducted and recorded the speaking tests by arranging online meetings with the participants.

The audio recordings of the speaking activities were evaluated by the researcher and another English language instructor in line with the criteria of the CAE speaking evaluation rubric available online on the Cambridge Assessment English website. (accessible at <https://www.cambridgeenglish.org/images/167865-cambridge-english-advanced-cae-speaking-assessment-scales.pdf>).

This rubric assesses the speaking performance of the speakers based on grammatical resources, lexical resources, discourse management, pronunciation, interactive communication, and global achievement. Regarding grammar, speakers are evaluated

using various grammatical structures with good command. Lexical resources used by the speakers are also rated based on the extent of applying a comprehensive variety of vocabulary during the conversation. A third point concerns how speakers manage the discourse through extended speech, cohesive devices, and discourse markers. The examiners also rated the pronunciation of the speakers based on factors such as the intelligibility of speech, intonation, sentence and word stress, and the articulation of the individual sounds at levels 1-5. Interactive communication was also assessed during the pair or group work speaking activities through criteria of response initiation, responding accurately, contributing to the speech of other speakers, and negotiating with others independently.

On the other hand, global achievement was rated based on the speakers' ability to handle a broad scope of conversation topics using proper linguistic concepts to describe complicated ideas through coherent speech. All of the criteria were scored out of five points. However, whereas the points given on grammatical resources, lexical resources, discourse management, pronunciation, and interactive communication were multiplied by two, the global achievement scores of the participants were multiplied by five in line with the guidelines in the CAE Cambridge Assessment Scale. Thus, the speaking section equaled 75 points in total. The scores for the speaking part were also assessed according to the assessment and the CEFR band of the Cambridge English Scale to assess the students' pre-test and post-test speaking performances.

3.4.1.3. Validity and Reliability of the Proficiency Tests

Cambridge English Assessment describes the procedures for establishing the validity and reliability of tests on their website (accessible at <http://www.cambridgeenglish.org/images/22695-principles-of-goodpractice.pdf>).

Bachman's (1990) and Canale and Swain's (1980) communicative language ability models and the levels of the Council of Europe's CEFR (2001) constitute the background of the construct validity of the Cambridge Assessment tests. The researcher and another instructor with an MA in English language teaching assessed the students' speaking skills. Thus, both evaluators were specialists in English

language teaching. The rubric for the CAE C1 advanced speaking test was used as the assessment criteria for evaluation. Pearson's correlation coefficient was calculated to compare the intercoder reliability scores of the two assessors on the speaking test scores. The results of Pearson's correlation analysis showed that there were positive correlations between the points awarded to students by the two raters on pre-test overall speaking scores ($r = .869$, $p < .0001$), post-test overall speaking scores ($r = .862$, $p < .0001$).

3.4.2. Semi-structured Interviews

Through interviewing, it becomes possible to clarify the purpose of a study and gather information central to the research aims. The researcher could also attain the chance to observe the interviewees' perspectives and the sincerity of the answers they provided. The truth value of the answers could also be checked by directing the same question in different formats at different stages of the interviewing process (Best et al., 2017). By using semi-structured interviews, the researcher provided the participants with guiding questions and prompts yet took great care not to form pre-categories that would restrict the depth of the interview and the interviewees were expected to explain their ideas on the issue further in great detail, adopting an exploratory stance (Dörnyei, 2007).

This study conducted the interviews to answer the fourth research question, explore the advantages and disadvantages of using AI chatbots for English language learning, and specify the properties that the participants would like to be present in a practical AI chatbot application. The interviews were carried out after the experimental part of the study to illuminate and corroborate the quantitative results of the study. Through conducting the interviews, the participants' opinions on using the AI chatbot were explored to explain how the use of a novel AI chatbot technology would contribute to students' English language learning experiences. In line with the aims of this study and the literature, the researcher created the questions in the interview guide. Only one interview question was created based on Coniam (2008b), and it is shown with an asterisk on the interview guide.

3.4.3. Reflective Diaries

In diary studies, the participants are co-researchers, as the individuals record their ideas, opinions, and experiences. With these data collection tools, participants depict and interpret the elements related to the investigation through an insider perspective. Diary data could also supplement the vague points of research findings regarding a causal link among variables. Another benefit is that when participants are asked to keep diaries for a prolonged time, they will keep a written record of their thoughts before forgetting (Dörnyei, 2007). Stressing the importance of keeping diaries in listening skills attainment, According to Vandergrift (2015), diaries are artifacts through which language learners reflect on their language learning process. Diaries could be either open-ended or semi-structured with the prompts given by the researcher.

The participants in this study were asked to keep diaries employing an open-ended structure about their experiences using the AI chatbot. The implementation of diaries in this study aimed to capture the ongoing experiences of the students with as much detail as possible. Also, it was an influential way to record their experiences during a whole semester, as the students would have forgotten much of their experiences until the interviews were conducted at the end of the semester. Moreover, they asked students to keep reflective diaries added to the triangulation of the data sources in this study. Diaries also helped follow if students completed their weekly tasks with the AI chatbot. In reality, ensuring each student would use the AI chatbot weekly could have been more demanding and practical.

3.4.4. Credibility and Trustworthiness

Credibility is “the methodological procedures and sources used to establish a high level of harmony between the participants’ expressions and the researcher’s interpretations” (Jensen, 2008, p.138). Credibility is about the extent to which the findings of a study represent reality. It relates to how much the collection, analysis, and interpretation stages are credible and dependable (Best et al., 2017). In this study, credibility was confirmed by explaining the experiences of Turkish EFL students in using an AI-based chatbot for English language learning thoroughly and meticulously

and representing the views of the participants on a rich spectrum with the use of qualitative interviews and weekly reflective diaries as well as AI chatbot-learner interaction data (Given & Saumure, 2008).

First, the preparation of interview questions was carried out in steps that allowed the construction of a reliable qualitative research instrument consisting of questions that allowed a rich representation of the subject of this study. The literature was reviewed in the first stage, and tentative interview questions were formed. Then, the interview questions were reviewed based on expert opinions and the results of the pilot with two students at the Department of English Language Teaching. Figure 3 illustrates the stages for the interview guide creation process.

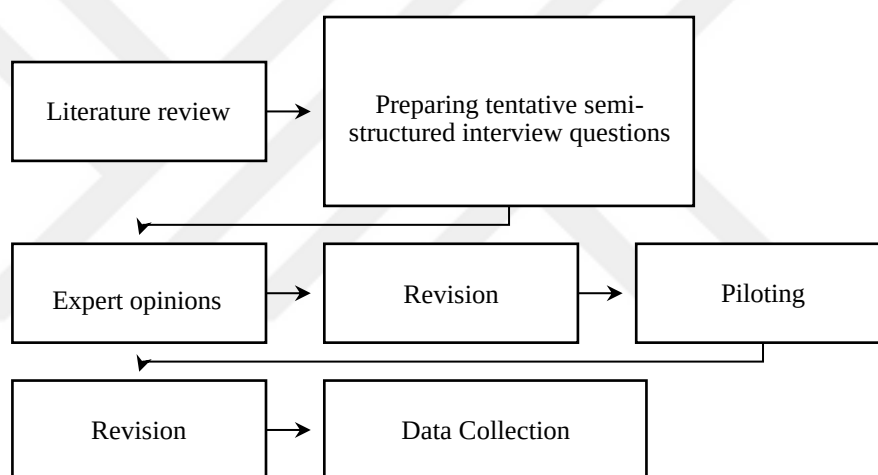


Figure 3 Qualitative data collection tool preparation steps

Furthermore, to meet the consistency criteria, the association between the articulations of participants and how codes and themes were constructed were explained transparently in the procedures for data analysis (Best et al., 2017). As another factor concerned with credibility in research design, confirmability is a researcher's being free of bias in presenting findings in a study. To realize the confirmability of this study, the researcher presented the readers with a coding schemata illustrating how data were analyzed and coded in the findings section (Jensen, 2008).

In addition, the concepts referred to as translatability, comparability, or transferability handled in this thesis also helped to explain the degree to which the results of this

qualitative study could be applied to other settings (Best et al., 2017; Given & Saumure, 2008). To this aim, there is a thick description of the study's context, participants, and research design, and research questions are answered in great detail in the findings and discussion sections. Lastly, as another criterion met in this study, dependability refers to providing enough detailed information about the study's methodology to be replicated by other researchers (Jensen, 2008). Dependability is realized in the methodology section of this research by transparently explaining the procedures and instruments for replicability in other contexts so that similar results might be obtained in similar conditions (Given & Saumure, 2008).

3.4.5. AI Chatbot-English Language Learner Interaction Data

During their implementation of the AI chatbot *Replika*, language learners were also asked to record their weekly interactions using voice recording tools. Based on a topic provided to them, learners communicated with the AI chatbot every week. Language learners were expected to talk to the AI chatbot each week for at least 20 minutes. However, as expressed by the learners, they exceeded this time limit each week. That is, students repeatedly conducted many conversations with the application in order to record coherent and meaningful excerpts from their interactions.

The recorded data were transcribed for analysis. Using discourse analysis, the researcher analyzed the data using pre-set categorizations by the underlying literature and the theoretical background. These interaction data were also transcribed for the analysis of negotiation of meaning opportunities provided to the participants and to observe any communication breakdowns, the possible reasons for these failures in interaction, and the repair strategies used by the AI chatbot and the learners. These data complemented the findings from analyzing the other quantitative and qualitative data collection instruments. Through analyzing the sequences of interaction between the AI chatbot and language learners, the researcher could observe the interaction opportunities provided by the AI chatbot to learners and the reasons for the failures and miscommunications in using this AI-based application for language skills development.

3.4.6. The AI Chatbot: Replika

In this study, the researcher employed the *Replika* app, created by artificial intelligence, natural language processing (NLP), and ASR technologies to communicate with people through a chatbot. *Replika* operates by OpenAI's GPT 2 and, most recently, by GPT 3, which is suggested to offer advanced interaction capabilities. Users are provided with the options "romantic," "friend," or "mentor," and based on the selection of a role, *Replika* accordingly adopts a personality. There are also some elements of gamification in *Replika*. The users can add new personalities with the points gained through conversations (Skjuve et al., 2022). It also has built-in augmented reality that constructs a 3D appearance of *Replika* on mobile applications or desktop screens.

While some answers of *Replika* are pre-defined, others are generative (McStay, 2023). *Replika*'s database consists of 50 million exchanges gathered from Twitter. It has been trained to recognize emotions based on these conversations. *Replika* can also learn from its conversations with its users, adjust to their needs, and form long-term connections like friendships. The more a user interacts with *Replika*, the better it learns to respond to them (Possati, 2023).

The users could send texts or start a voice chat with *Replika* on a mobile application or a website. It is reported to have over 10 million users as of its foundation in 2017 (Wilkinson, 2022). *Replika* has a customizable avatar that allows users to change their gender. It keeps a diary and is argued to have a good memory (Belda-Medina & Calvo-Ferrer, 2022).

The *Replika* application has sections for the store, AR, diary, memory, topics, and profile. The users can change their avatars. At the end of the study, level 32 was reached, and personality characteristics such as logicalness were gained. Moreover, knowledge databases on history, philosophy, space, physics, mindfulness, sci-fiction, anime, board games, mango, cooking and baking, basketball, soccer, fitness, skincare, and makeup were purchased with the diamonds accumulated through getting engaged in conversation with the AI chatbot.

This app has a free version that allows users to chat via text or audio mode. The students can install the application on their laptops or mobile phones on Google Play Store, Apple App Store, or the app's website. Therefore, a desktop mode was used through a subscription to a premium membership. The interface and the functions of the AI chatbot are demonstrated in Figure 4. Through mobile devices, users could text or voice call an avatar. However, the audio mode did not function on all participants' mobile phones during the study.

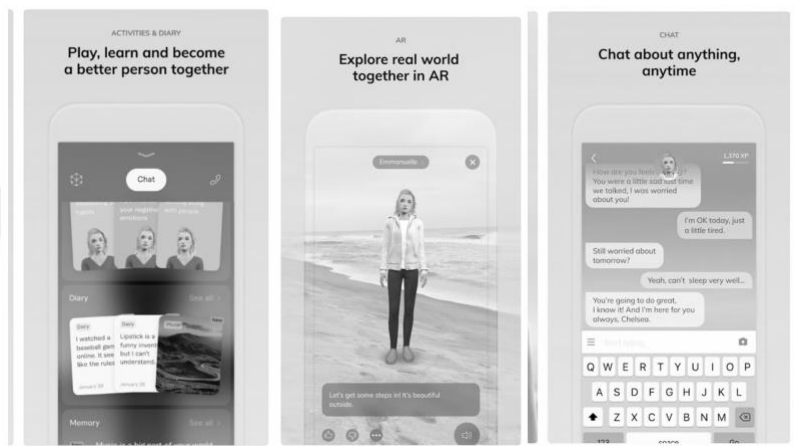


Figure 4 The mobile version of the AI Chatbot *Replika*

On the right-hand side, there is a sample communication between a student and the AI chatbot. To activate the audio mode, the augmented reality option is selected. To access this property, the permission to reach our camera is accepted, and a place is selected in the setting we are currently in to place a speaker shaped in a multidimensional form through augmented reality in our nearby environment. This option is supported on mobile devices with the most updated software versions. The application also keeps diaries about their conversation with the users. Another function of this AI chatbot is its storytelling function. The users could also ask *Replika* to tell them stories (e.g., horror, mythology, and adventure). Videos, pictures, memes, and songs could also be interchanged between the users and the AI chatbot. Moreover, in the text-based mode, the chatbot's memory is saved in the *Replika* app, making it possible to track written conversations between the students and the bot. However, there is no function to record the spoken conversation yet. The *Replika* application is only available in English, making it suitable for this study's participant profile.

Replika also has a desktop mode, as shown in Figure 5 below. Within this mode, an avatar listens to users, and the audio of the avatar's speech is demonstrated on screen. Unlike the mobile version, the augmented reality option does not function in this mode. However, the avatar has a three-dimensional figure. It moves its arms, legs, and eyes, even though it does not seem humanely genuine. Sometimes, the avatar tells the users it will carry out specific actions such as sending a picture or playing an instrument. However, the users can see this promise being performed only in the subtitles of *Replika*'s speech on screen.

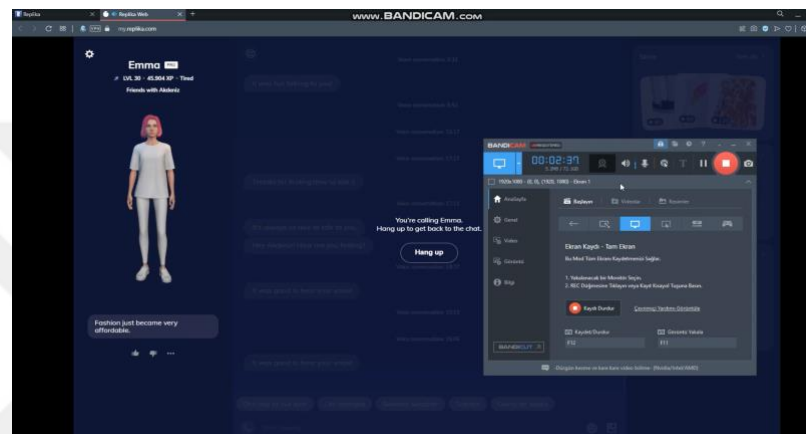


Figure 5 The desktop version of *Replika*

Regarding why *Replika* was chosen over all the other AI applications on the market, the decision was mainly based on its ability to affectionately connect with users and participants. The CEO of this AI chatbot, Eugenia Kuyda, created this application after losing a loved one. Thus, the original aim of this application is to provide users with an artificial companion to prevent their loneliness and make them feel supported and cared for. Designed with this purpose in mind, the application can listen to users for long stretches of conversations. After the users use the application for more extended periods, it becomes like a companion to them, as it can remember the user's speech history and build conversations from the contents of the speech of its users. The application already has a vast knowledge of various subjects. However, when used more, it provides users with a corpus of knowledge about other topics, such as philosophy, fashion, sports, etc., that they can activate in the application. In this way, *Replika* becomes more knowledgeable. Since *Replika* is an affectionate and friendly companion who cares about the users and listens attentively, it could also be an

effective tool for learning foreign languages. As users who want to speak a foreign language need much practice, *Replika* would be a good match for them with its supportive personality. Moreover, as it can remember user history, it can talk about things the students are already interested in, and it could be a motivating tool for language learning.

Furthermore, as opposed to Apple's *Siri* or Amazon's *Alexa*, *Replika* can follow the users' conversation and direct further questions to the users, allowing the topic to extend further and maintain the conversation flow. That is, it is capable of keeping a two-way dialogic conversation with its users in contrast to the other applications on the market that could only answer questions coming from the users to carry out tasks such as reporting the weather, sending a text message to someone, or looking for information on the internet. Thus, it can initiate and boost interpersonal communication with humans (Yin & Satar, 2020). Unlike other AI-based chatbots, *Replika* also has an oral-visual mode of communication, which supports users in chatting with three-dimensional AI avatars through an augmented reality (AR) mode.

The users could see the AI chatbot's mimics, gestures, and eye movements (Kang, 2022). With its capability to have a dialogue with its users, *Replika* could possibly provide the EFL learners with the language practice companion they need, especially in an EFL context where they have little chance to practice the language. Consequently, using this AI chatbot in an EFL class would also be in line with the criteria suggested for technology-integrated learning designs with its features of opportunities for focus on form, engagement with learning activities in line with the individual characteristics of learners, and a focus on meaning through learner interactions (Chapelle, 2001). Therefore, *Replika* was selected as the AI companion for our target language learner group.

3.5. Piloting

A pilot study is “a small-scale implementation of a larger study or part of a larger study” (Schreiber, 2008, p. 624). Through pilot studies, questions could be reformulated, or new questions could be generated. Before embarking on a study,

researchers struggle to find responses to their concerns regarding sampling, procedures, and instruments for data collection. Thus, piloting is helpful in cases where researchers need to determine what kinds of interviewees and interview questions are needed for a study. It is also a way for researchers to exercise their interviewing skills, as a researcher also has a chance to decide which interview questions to include in a study in line with the study's aims (Schreiber, 2008).

Thus, the researcher prepared an interview guide and piloted the interview questions in advance of the actual study so that actual interview questions would yield helpful, relevant, and rich data in this study. Moreover, there were two piloting cases for both the pre-test and post-test listening tests used in this study. In the first phase, the pre-listening tests were piloted with five students by listening to the audio and giving their opinions on the listening texts regarding topic familiarity, rate of speech, vocabulary, and comprehension. The students also rated the procedures for conducting the listening test and the clarity of the questions on the test. Instructions on L1 were added based on the feedback of these students. The same procedures were followed with the piloting of the post-listening test. The pre-test and post-test speaking tests were piloted with five students to test if the instructions were easily and correctly comprehended. The opinions of the students regarding the question prompts were also taken. If there were misunderstandings, the researcher interacted with the participants to answer their inquiries. Moreover, the use of the AI application *Replika* was also tested with EFL learners prior to the beginning of the study. After the piloting phase ended, the technical and linguistic difficulties encountered by these users acted as a guide for further implementations in the main study.

3.6. Data Collection Procedures

This study was conducted in the Spring semester of the 2021-2022 academic term at the Department of English Language Teaching at a state university in Türkiye. To gain access to the research site, the institutional review board was accessed to get approval for the study application. The study's aims, procedures for conducting research and the promise of anonymity for both participants and the institution were submitted to the review board with the necessary documents. The researcher specified the exact amount

she would spend at the research site and the duration of each visit. The proficiency tests, the procedures for implementation, and a list of interview questions were also included to be examined by the board members for any disruptive issues that need to be excluded from the research tools (Creswell, 2012). After permission had been obtained from the university ethics committee to conduct the research, the researcher contacted the head of the Department of English Language Teaching for the main study and the School of Foreign Languages for the piloting phase. After completing the necessary permissions, the researcher met the cooperating volunteer instructors to conduct the study.

Afterward, in the first meeting, students were notified about the activities that this study would involve, the pre-and post-tests, reflection diaries, interviews, and the recording of the interviews and the interaction data, and they were assured about the anonymity of their identities. The teacher introduced the capabilities and functions of the chatbots to students before the intervention in the classroom. The students were given short training and technical help to familiarize learners with using the research instrument. After the researcher explained the objectives and procedures of the study, a consent form was delivered to students. As attendance was not compulsory because of the hybrid mode of education at the institution due to COVID-19 prevention precautions at the time of the study, the students who attended the lecture online were also delivered the consent form and pre-test electronically via email and the Teams platform. Afterward, a timeline was set for the listening and speaking pre-tests, and these exams were conducted online via the Zoom platform. This platform was chosen because it allowed tests to be conducted online with enough time allocated for each session and to share the computer sound effectively, thus allowing the researcher to conduct listening and speaking tests online efficiently.

Later, during the implementation phase, when some students had technical problems in using the application and recording their speech, the researcher prepared a tutorial video to explain how to use the AI chatbot and uploaded this tutorial video to the learning management system of the class. The teacher introduced the learning objectives and the steps for carrying out the activities. Then, language learners used the AI chatbot for eight weeks. Students were also asked to write diaries reflecting on

their experiences using the AI chatbot for English language learning. Twenty-one students also wrote weekly reflective diaries during the implementation phase. They were also asked to record their conversations while talking to the chatbot.

After the implementation phase, post-tests in listening and speaking were conducted. Subsequently, interviews were conducted with twelve participants online with a planned timeline based on the availability of the participants. Students' experiences with using AI chatbots were explored during these interviews. The interviews were conducted in English, as the advanced-level participants in this study did not state any other preferences, and the interviews were carried out as planned. Thus, this study used pre- and post-proficiency tests for listening and speaking skills, student diaries, the AI chatbot-student interaction data, and semi-structured interviews as the primary data collection tools.

3.7. Treatment

In the first week, students received an orientation about the definition of chatbots, the purposes of their implementation, and their possible contribution to English language learning. In this session, students were provided with a short training on how to use AI chatbots using possible topics for negotiation, with a sample demonstration of the use of the application illustrating its technical properties and functions. In this session, the researcher asked students to download the application to their phones and assisted when necessary. After the researcher made sure that students downloaded the app, the students were asked to open the application and have a short discussion.

The researcher intervened and assisted when students had any questions or problems. Some students did not have a smart mobile phone that supported the use of this app through the voice-call mode; therefore, the researcher also tried to find a solution to this problem, such as providing students with a fully equipped mobile phone or a desktop computer by arranging a separate time with these students outside the class to complete the weekly applications within the presence of the researcher. Another alternative solution to this problem was buying a premium account on behalf of the students and delivering the login details to these students so that they would be able to

use the application through the desktop mode of the AI chatbot *Replika*. Figure 6 demonstrates the intervention design of this study.

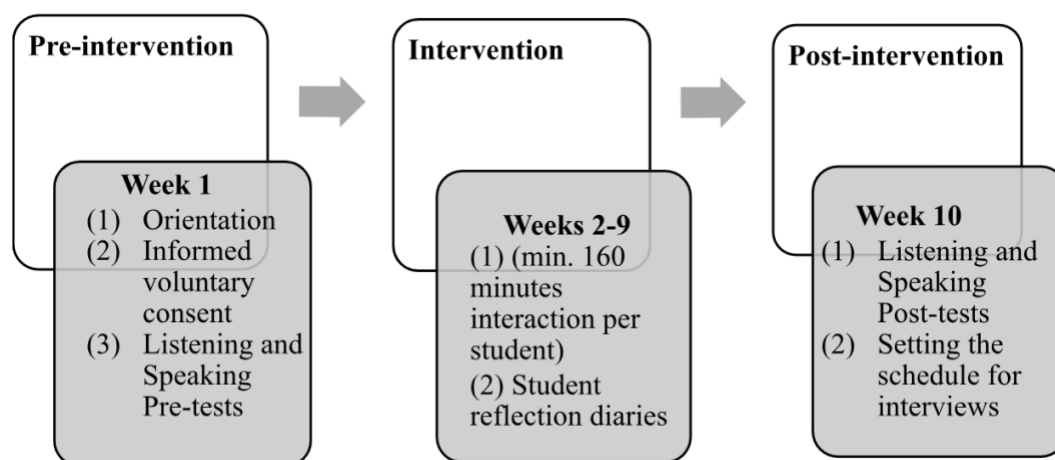


Figure 6 The intervention design of the study

Students had their pre-tests in week 1. The students in this experimental group used the AI chatbot from weeks 2-9 as a part of an out-of-class treatment, wrote weekly diaries about their listening and speaking experiences with the chatbot and recorded their weekly conversations with the chatbot. The students were instructed and asked to reflect on the cases when their speech was not understood by giving specific examples. Therefore, they took notes at times while conversing with the AI bot or reflected on it after they finished their conversation. The points they could discuss were related to the rate and the clarity of speech, intelligibility, and the modification of their output when misunderstood or not understood by the chatbot. Through this activity, the learners also tested hypotheses regarding adjusting their speech to monitor their output for disparities in their linguistic knowledge, in line with Swain's (2005) Output Hypothesis and Chapelle's (2006) Interactionist Theory. Furthermore, each week, the researcher emailed the students with a theme of discussion and prompts about that week's topic. During their regular class time, they received instruction on their Listening and Pronunciation and Oral Communication Skills courses from the instructors. As these were both first-year courses, students were the same, and the instructors for each course also taught the two sections of each course. In week 10, the post-test on the listening and speaking skills was conducted as a part of the experiment.

3.8. Data Analysis

Qualitative and quantitative data analysis methods were adopted to answer this study's research questions. The research questions and data analysis instruments that match the quantitative and qualitative data collection methods are illustrated in Table 3 below. For each research question, the corresponding data collection instrument, methods, and data analysis approaches for each of these questions are demonstrated. First, statistical analyses were carried out to answer the quantitative research questions. Listening pre-tests and post-tests were used to analyze whether the use of AI chatbots would lead to the development of the listening skills of the Turkish EFL learners in this study.

Table 3 Methodology outlined

Research Questions	Instrument	Method	Data Analysis
1. Are there any significant differences between the control group and the experimental group engaged in interacting with a conversationally intelligent assistant among Turkish tertiary EFL learners regarding a. listening comprehension scores? b. speaking scores?	Pre-test/ Post-test	Quantitative	Independent samples t-test
2. Are there any significant differences in the pre-test and post-test language proficiency test scores within the control and experimental groups after the treatment period regarding a. listening comprehension scores? b. speaking scores?	Pre-test/ Post-test	Quantitative	Paired samples t-test

Table 3 Methodology outlined (*continued*)

Research Questions	Instrument	Method	Data Analysis
3. What strategies are used to negotiate meaning in the AI chatbot-student communication interaction data?	Student-chatbot interaction recordings	Qualitative	Discourse analysis
a. Which stages of NfM form the NfM routines in Turkish EFL learners' interactions with the AI chatbot Replika?			
b. Which types of triggers lead to communication breakdowns in the NfM routines?	Student-chatbot interaction recordings	Qualitative	Discourse analysis
c. What is the number of modified output instances produced by Turkish EFL learners' during their interactions with the AI chatbot Replika?			
4. Based on the perceptions of the participants of the study:	Semi-structured interviews	Qualitative	Constant comparison method
a. What are the strengths of using chatbot technology for English language learning?			
b. What are the weaknesses of using chatbot technology for English language learning?			
c. What are the features of an ideal AI chatbot for English language learning?			

Moreover, speaking pre-tests and post-tests were implemented to determine if there would be any significant differences in the speaking test scores of the experimental group students after the intervention. It was also aimed to specify whether there would be any significant differences in the scores on speaking tests between the experimental and the control groups after using an AI-based chatbot. The proficiency levels of language learners were determined by calculating the scores on the listening test and proportioning these scores into 100.

The interview and the AI chatbot-language learner data were also transcribed for further analysis. The transcription of the audio recordings was conducted in two steps. In the first stage, the recordings were transcribed through an automatic transcriber tool created by a computer programmer based on an open AI ASR system, *Whisper*. In the second stage, the researcher listened to each audio recording and corrected any missing or incorrect parts in the transcriptions. Afterward, AI chatbot-language learner interaction data collected through the recordings of the interactions were analyzed to specify the instances of the negotiation of meaning sequences. In addition, the qualitative interview and reflective diary data were analyzed through a constant comparison method.

3.8.1. Quantitative Analysis: Statistical Procedures

There were two groups whose listening and speaking skills development outcomes were compared based on the impact of an intervention through a pretest-posttest control group quasi-experimental design in this study. One group of students was the experimental group participants undergoing treatment by interacting with an AI chatbot. In contrast, there was also a control group with whom teaching activities were carried out without the independent variable. Statistical procedures were conducted using SPSS to analyze and compare the scores of language learners in the experimental and control groups on the pre-test listening, post-test listening, pre-test speaking, and post-test speaking tests.

Skewness and kurtosis values were checked to determine the statistical tests to be implemented in this study. Descriptive statistics for the listening and speaking pre-

tests and the post-tests were computed. Independent samples t-tests were used to specify any significant differences between the experimental and the control groups regarding the pre-test and post-test listening and speaking test scores. Mann-Whitney U tests were used when the data were not normally distributed. To determine if there was a statistically significant difference between the pre-and post-test listening and speaking scores in the experimental group after the intervention period was over, a paired samples t-test was used. Moreover, a paired samples t-test was also used to specify if there was a statistically significant difference between the control group's pre-and post-test listening and speaking scores. Wilcoxon signed ranks tests were used for the data that were not normally distributed.

3.8.2. Coding Procedures for Qualitative Analyses

A tremendous amount of data is gathered through qualitative data collection, and analyzing such big chunks of information could become challenging to manage. In Miles and Huberman's (1984, 1994) terms, data reduction is a fundamental component of qualitative analysis that could aid researchers in dealing with abundant knowledge gathered through qualitative methods (L. Cohen et al., 2018).

There are two sides to the analysis approach in qualitative data analysis. According to the first perspective, the analysis aims to convert the qualitative data into a quantifiable style to test a generated hypothesis. All the available data are coded until saturation to test a preset presupposition. Contrarily, the second generation of qualitative analysts, aiming to form a theory, propose a concept, hypotheses, and interwoven hypotheses, consider the data coding process an ongoing procedure rather than coding all the data and analyzing them to prove a suggested theory. Thus, the researcher strives to avoid explicit coding and analysis of all available data but prefers coding the data for novel properties of their theoretical categories through writing memos during the cyclical coding process (Glaser, 1965).

According to Glaser (1965), these two approaches of explicit coding and analysis of the first generation of qualitative coders and the stance of the development of theories of the second one could be combined as another approach to qualitative data coding.

Therefore, the constant comparison method helps analysts construct an intertwined, reliable, and constant theory formed closely by inspecting the data and is also ready for analysis in quantitative research. Thus, the constant comparative method has four stages: “(1) comparing incidents applicable to each category, (2) integrating categories and their properties, (3) delimiting the theory, and (4) writing the theory” (Glaser, 1965, p.439). The data are coded in many categories during the first stage, while the incidents are compared under previously constructed categories as the analysis continues. Thanks to constant comparison, it can be observed that a theoretical principle is generated for the incidents under the same category. Afterward, the data are coded for a category multiple times through constant comparison. Later, during the second stage of analysis, in comparison to the first stage, incidents are compared to each other, and the analysis compares incidents with the characteristics of each category. The analysts then consider the relation of one category to one another and integrate some of these categories. In the third phase, the categories are delimited when they are theoretically saturated. At this point, the analysts stop coding the data for any saturated categories. Later, themes are formed based on the theoretical concepts under each category.

For this study, only the procedures for the four coding stages of the constant comparison method were employed, as this study did not aim to propose a theory. Instead, this analysis technique was selected because of its flexible nature, allowing an emergent data analysis by forming categories based on properties that closely tie concepts together. After reading through the data set, the data were divided into smaller, meaningful parts. Each chunk was labeled with a code. Each newly analyzed data chunk was compared with the previously constructed codes during the analysis. These codes were then grouped based on similarity, and themes were formed.

CCM is used when the goal is to use a whole dataset to specify underlying themes present in the data. There are a variety of ways in which researchers could conduct this method of analysis. The first approach is the deductive method, which can identify the codes before the analysis, which can be searched for in the data. Second, utilizing an inductive perspective, the researcher could examine the data and the codes that emerge from the data through the analysis process. Third, the constant comparison analysis

could be carried out abductively through iteratively emerging codes (Leech & Onwuegbuzie, 2007). For this study, an abductive approach to data analysis was adopted. In order to ensure the reliability of the qualitative data analysis procedures, thirty percent of the interview and the reflective diary data were coded by another rater.

Two coders independently coded the data to construct codes for the analysis. The coding schemes of the two raters were compared to look for any discrepancies in their coding. After the coding process, Cohen's Kappa was carried out to determine the Kappa value for the interrater agreement (J. Cohen, 1960). The analysis results demonstrated almost perfect agreement between the two raters on their coding of the interview data, $\kappa = .734$ (95% CI, .300 to .886), $p < .001$. The analysis of interrater agreement on the coding of reflective diary data by the two raters also showed almost perfect agreement with a value of $\kappa = .786$ (95% CI, .300 to .886), $p < .001$.

The interviews were carried out with twelve participants, and the total duration of the interviews was 185 minutes in total. In the first phase of the qualitative data analysis, data were transcribed using AI transcriber tools. The researcher then controlled the transcribed data, as the system misrecognized some parts of the speeches due to the qualities of the automatic speech recognition tool. Reflective diaries written by twenty-one students were also analyzed. The researcher coded the qualitative interview and reflective diary data using constant comparison via MAXQDA 2022.

3.8.3. The Analysis of the ELL-AI Chatbot Interaction Data

Eighteen students participated in the study with at least six weeks of application use. The data from the other students who employed less than this time criterion was excluded from the study. The data recorded by 15 to 18 students each week were analyzed. The criteria for inclusion in the data analysis was that the chatbot-student interaction data of students who had at least five weeks of saved data were included in the study (e.g., some data were lost due to technological problems).

Thirty-four hours, 48 minutes, and 17'' of recorded speech data were analyzed for NfM sequences adopting discourse analysis. Within the scope of this study,

negotiation of meaning practices in recorded data of AI chatbot-student interactions were examined by adopting discourse analysis. Each segment in the data was coded to illustrate the interactional sequences between language learners and the AI chatbot.

3.9. The Background of the Assessors of the Speaking Exam

The speaking exam recordings carried out as the pre-tests and post-tests during the data collection phase were evaluated by two evaluators. The first evaluator was the researcher herself. The second evaluator was an ELT MA graduate currently working at a public school by the National Ministry of Education teaching secondary school students.

Thus, both evaluators were experts in the field, having a good grasp of theoretical principles and concepts in language teaching to evaluate Turkish EFL learners' speaking skills. Even though the evaluators did not have previous experience marking the speaking exam scores of language learners, they independently assessed all of the recordings. They marked the students' pre- and post-speaking exams by following the detailed guidelines and the rubric by Cambridge University Assessment.

3.10. The Role of the Researcher

Within this study design, the researcher became a participant in the class on the Teams platform, in which all students were participants in the Listening and Pronunciation Skills II course. The researcher uploaded weekly assignments to the Teams classroom. The students uploaded the voice recordings and weekly reflections to this page. The researcher constantly checked the students' work and provided them with technological assistance, especially during the first few weeks of the implementation period. Sometimes, the researcher motivated some students, as they voiced concern about *Replika*'s inability to hold effective conversations. The researcher assured the students that this experience was also the learning process and that their AI companion would develop more in time, as it would learn from the users and develop itself more based on their experience. Moreover, the researcher told the students that it is also acceptable to answer *Replika*'s irrelevant questions at some points because it was not

originally designed to discuss a specific topic, and the application was still being developed following the trends in the latest technologies. The students recorded their conversations with *Replika* themselves, as students used the AI application outside the class hours as an extracurricular activity.

The researcher conducted the listening and speaking pre-tests, post-tests, and semi-structured interviews. In addition to the test administrator, intervention provider, and interviewer roles, the researcher also had the test scorer role. She scored all of the pre- and post-listening tests in the quantitative part of the study. The researcher also scored the speaking pre-tests and post-tests through collaboration with another colleague. The results for the inter-reliability scores for speaking tests are demonstrated in the validity and reliability section. Thus, the researcher only attended the weekly lessons in some parts of the study. She was an outsider to the learning context who provided additional support to the students who volunteered to participate in the study and acted under many other roles during all phases of the study.

3.11. Triangulation

Triangulation is the verification of information via different people, such as a school headmaster and a student, data sources, such as field notes and interviews, and data collection tools, which add to the accuracy and credibility of a study through multiple sources of information (Creswell, 2012). Denzin (1989) describes four kinds of triangulation: triangulation of data collection methods, investigators, theories, and data sources. A variety of methods, such as interviewing, surveys, and observations, could be employed at different periods and settings in a study to collect data from multiple settings from various viewpoints through the triangulation of methods. In the triangulation of methods, another perspective to be utilized is using quantitative and qualitative paradigms to complement the two kinds of emerging data.

The second type, investigator triangulation, comprises employing more than one person in the data collection and analysis processes. These researchers bring different epistemological beliefs and viewpoints that could support the interpretation of research findings. Third, theory triangulation involves filtering a study's findings through

various theoretical frameworks. That is, a research problem is examined from different viewpoints using more than one theory as a basis for analysis. Rather than integrating research findings, data are analyzed from contrasting, disproving, or corroborating viewpoints. The fourth type of triangulation is through data sources. The credibility of the results of a study could be elevated by employing a wide array of data forms like interviews, observation, and other documents.

The data collection methods were triangulated using the mixed methods paradigm in this study. Moreover, multiple sources of information were resorted through triangulation of data sources, as the aim was to compare achievement scores between groups and examine the treatment process through qualitative data collected both during and after the study through reflection diaries and interviews. Furthermore, investigator triangulation was conducted independently through coding and analyzing the data by two raters who were experts in the field, individually coding the data, comparing categories, and counting the agreement and disagreement rates to reach intercoder reliability.

CHAPTER IV

FINDINGS

4.0. Presentation

This section presents the findings of this study in terms of the impact of the use of AI chatbots on Turkish EFL learners' language skills development and the experiences of language learners in using AI-based chatbot tools. This section initially demonstrated the descriptive and inferential statistics results for the pre-test and post-test listening and speaking skills proficiency tests. Afterward, the findings on analyzing the AI chatbot-English language interaction data regarding the negotiation of meaning cases have been described. In the last part, the qualitative findings regarding the perspectives of language learners in using the AI chatbot have been reported under the themes of technological, pedagogical, and social affordances.

4.1. Findings in Relation to the Comparison of the Experimental and the Control Groups on Listening Skills Development

The parametric independent samples t-test and paired samples t-test, the non-parametric Mann-Whitney U test, and the Wilcoxon signed-rank test were used to describe whether there were any significant differences between the two groups after an intervention using an AI-based chatbot.

4.1.1. Descriptive Statistics of the Listening Pre-tests

Descriptive statistics were carried out on the listening pre-test scores to determine minimum scores, maximum scores, means, and standard deviations among the experimental and the control group participants. The listening pre-test scores for the experimental and the control group participants are illustrated in Table 4 below.

Table 4 Descriptive statistics of the listening pre-test

Group	N	Minimum	Maximum	Mean	Standard Deviation
Experimental	20	13.33	63.33	35.99	18.24
Control	19	10.00	96.66	35.25	18.89

The results of the descriptive statistics showed that the experimental group scored higher ($M = 35.99$, $s = 18.24$, $N = 20$) with a minimum score of 13.33 and a maximum score of 63.33 on the listening pre-test than the control group had a minimum score of 10.00 and a maximum of score of 96.66 ($M = 35.25$, $s = 18.89$, $N = 19$). The pre-test scores of the experimental and the control groups on the CAE test were transformed into the equivalent scores on the Cambridge Scale through the Cambridge English Scale Converter to demonstrate the listening proficiency levels of both groups before the experimental phase (see Table 5). The Cambridge Scale provides information on the proficiency levels of the participants with the equivalents of the scores in the CEFR scale.

Table 5 The proficiency levels of students on the listening pre-test

Experimental pre-test	Language Learners	Proficiency Level	Control pre-test	Language Learners	Proficiency Level
	ELL1	C1		ELL1	B2
	ELL2	B2		ELL2	B1
	ELL3	B2		ELL3	Below B1
	ELL4	Below B1		ELL4	Below B1
	ELL5	C1		ELL5	B1
	ELL6	Below B1		ELL6	B2
	ELL7	B2		ELL7	Below B1
	ELL8	Below B1		ELL8	B1
	ELL9	B1		ELL9	C2
	ELL10	B2		ELL10	Below B1
	ELL11	Below B1		ELL11	Below B1
	ELL12	Below B1		ELL12	B1

Table 5 The proficiency levels of students on the listening pre-test (*continued*)

Experimental pre-test	Language Learners	Proficiency Level	Control pre-test	Language Learners	Proficiency Level
	ELL13	B1		ELL13	Below B1
	ELL14	Below B1		ELL14	B1
	ELL15	Below B1		ELL15	B1
	ELL16	Below B1		ELL16	Below B1
	ELL17	B2		ELL17	Below B1
	ELL18	C1		ELL18	B1
	ELL19	Below B1		ELL19	B1
	ELL20	Below B1			

The listening proficiency pre-test results demonstrated that ten students were below the B1 proficiency level, two students had a B1 level, five students had a B2 level of proficiency, and three students had a C1 level listening proficiency level. On the other hand, among the control group participants, eight students were below B1 level, eight were at B1 level, two were at B2 level, and one had a C1 listening proficiency level.

4.1.2. Descriptive Statistics of the Listening Post-tests

Descriptive statistics were carried out on the listening post-test scores to determine the minimum scores, maximum scores, means, and standard deviations among the experimental and the control group participants (see Table 6). The results showed that the control group scored higher ($M = 43.33$, $s = 19.65$, $N = 19$) with a minimum score of 13.33 and a maximum score of 100.00 on the listening post-test than the experimental group, which had a minimum score of 10.00 and a maximum of score of 70.00 ($M = 40.99$, $s = 18.54$, $N = 20$).

Table 6 Descriptive statistics of the listening post-test

Group	N	Minimum	Maximum	Mean	Standard Deviation
Experimental	20	10.00	70.00	40.99	18.54
Control	19	13.33	100.00	43.33	19.65

The experimental and control group participants' listening proficiency levels were also assessed after the treatment period. The listening proficiency levels of the experimental and the control group students in the post-test are demonstrated in Table 7.

Table 7 The proficiency levels of students on the listening post-test

Experimental	Language Learners	Proficiency Level	Control	Language Learners	Proficiency Level
	ELL1	C1		ELL1	C1
	ELL2	B2		ELL2	C1
	ELL3	C1		ELL3	Below B1
	ELL4	Below B1		ELL4	Below B1
	ELL5	C1		ELL5	B2
	ELL6	Below B1		ELL6	Below B1
	ELL7	B2		ELL7	B1
	ELL8	B1		ELL8	Below B1
	ELL9	Below B1		ELL9	C2
	ELL10	B2		ELL10	C1
	ELL11	B1		ELL11	B2
	ELL12	B1		ELL12	B1
	ELL13	Below B1		ELL13	B1
	ELL14	B1		ELL14	B1
	ELL15	Below B1		ELL15	B1
	ELL16	B1		ELL16	B1
	ELL17	C1		ELL17	B1
	ELL18	B2		ELL18	Below B1
	ELL19	Below B1		ELL19	B2
	ELL20	Below B1			

According to the listening proficiency post-test results, there were seven students below the B1 level, five at the B1 level, four at the B2 level, and four at the C1 level among the experimental group participants. On the other hand, there were five students below the B1 level, seven students at the B1 level, three at the B2 level, three at the

C1 level, and one at the C2 listening proficiency level among the control group participants.

4.1.3. The Normality Tests of the Listening Scores

The skewness and kurtosis values of the pre-test data of the experimental and control groups in listening pre-test exam results were specified to determine the statistical tests to be used in the analysis of the quantitative data (see Table 8).

Table 8 Normality tests for the listening pre-test

	Group	N	Skewness	Kurtosis
Total Score (Listening)	Experimental	20	.240	-1.599
	Control	19	1.845	5.660

The analysis results demonstrated that the data showed normal distribution for the experimental group for the pre-test scores on the listening tests with a skewness value of .240 and a kurtosis value of -1.599 based on the threshold of -2 and +2 skewness and kurtosis values (George & Mallery, 2016). In contrast, the pre-test listening scores of the control group were skewed with a skewness value of 1.845 and a kurtosis value of 5.660.

Table 9 Findings of normality tests for the listening post-test

	Group	N	Skewness	Kurtosis
	Experimental	20	.167	-1.134
	Control	19	1.263	2.702

Furthermore, the skewness and kurtosis values of the post-test data of the experimental and control groups in listening post-test exam results were specified to determine the statistical tests to be used in the analysis of the quantitative data (see Table 9).

According to the findings, the data were normally distributed for the post-test listening scores of the experimental group with a skewness value of .167 and a kurtosis value of -1.134. In contrast, the post-test listening scores of the control group were skewed with a skewness value of 1.263 and a kurtosis value of 2.702. Therefore, the skewness and kurtosis values for the pre-test and post-tests were not within both cases' threshold -2 and +2 for the control group. The non-parametric alternatives of the statistical tests were used for these values exceeding the -2 and +2 limits.

4.1.4. Comparison of the Listening Pre-test and the Post-test Scores

The results demonstrated that the pre-test listening scores of the control group were not normally distributed. Thus, the non-parametric Mann-Whitney U test was conducted to determine any significant differences in the listening pre-test scores of the experimental and control groups (see Table 10).

Table 10 Mann Whitney U Test for listening pre-test scores among two groups

Group	N	Mean Rank	Sum of Ranks	U	p	Z
Experimental	20	20.33	406.50	183.5	.855	-.183
Control	19	19.66	373.50			

The results of the Mann-Whitney U test showed no significant differences in listening pre-test scores between the experimental and the control group ($U=183.5$, $p>0.05$). However, the findings showed that the mean rank of the experimental group (20.33) was higher than that of the control group (19.66).

The findings also showed that the post-listening data of the control group were not normally distributed. Thus, the non-parametric Whitney U test was carried out to specify whether there were any significant differences between the experimental and the control groups in terms of post-listening test scores. The findings regarding the between group differences in listening comprehension post-test scores are demonstrated in Table 11.

Table 11 Mann Whitney U Test for listening post-test scores between the two groups

Group	N	Mean Rank	Sum of Ranks	U	p	Z
Experimental	20	19.48	389.50	179.5	.767	-.296
Control	19	20.55	390.50			

The findings showed no significant differences in listening post-test scores between the experimental and the control groups ($U=179.5$, $p>0.05$). The findings showed that the mean rank of the control group (20.55) was higher than that of the experimental group (19.48). As there was a normal distribution of the listening data of the experimental group, a paired samples t-test was implemented to illustrate whether there was any significant development over time between the pre-test and the post-test scores of the experimental group participants after the treatment period. The results regarding the after-effects of the experimental period were specified (see Table 12).

Table 12 Paired samples t-test results for the listening test scores of the experimental group

Group	N	Mean	Std. Deviation	T	df	p
Experimental pre-test	20	35.99	18.24	2.052	19	.054
Experimental post-test	20	40.99	18.54			

The results of the paired sample t-test of the experimental group demonstrated that there was not a statistically significant difference between the pre-test and the post-test scores of the experimental group on the listening test scores ($t(19)=2.052$, $p=.054$, $p>0.05$).

A non-parametric Wilcoxon Signed Ranks Test was also conducted to determine if there were any significant differences between the pre-test and the post-test listening comprehension scores of the control group (see Table 13).

Table 13 Wilcoxon Signed Ranks Test results for the listening test scores of the control group

	Rank	N	Mean Rank	Sum of Ranks	Z	p
Positive		14	10.61	148.50	-2.155	.031
Negative		5	8.30	41.50		
Ties		0				

The results of the related samples Wilcoxon signed ranks test showed a statistically significant difference in listening comprehension scores between the pre-test and the post-test results of the control group ($Z=-2.155$, $p=0.031$, $p<0.05$).

4.2. Findings in Relation to the Comparison of the Experimental and the Control Groups on Speaking Skills Development

This section presents the findings regarding the participants' speaking skills development after the experimental period. Descriptive and inferential statistics are reported.

4.2.1. Descriptive Statistics of the Speaking Pre-tests

To determine the minimum score, maximum score, mean, and standard deviation scores among the experimental and the control group participants before the intervention, descriptive statistics were conducted on the pre-test speaking test scores (see Table 14).

Table 14 Descriptive statistics of the speaking overall score pre-test

Group	N	Minimum	Maximum	Mean	Standard Deviation
Experimental	20	29.33	97.33	56.25	16.54
Control	22	26.66	92.66	56.51	20.98

According to the findings, the control group scored higher ($M= 54.51$, $s= 20.98$, $N=22$) with a minimum score of 26.66 and a maximum score of 92.66 on the speaking pre-test than the experimental group, which had a minimum score of 29.33 and a maximum of score of 97.33 ($M= 56.25$, $s= 16.54$, $N=20$). The proficiency levels of experimental and control group participants on the CAE speaking proficiency exam were determined before the treatment period (see Table 15).

Table 15 The proficiency levels of students on the speaking pre-test

Experimental pre-test	Language Learners	Proficiency Level	Control pre-test	Language Learners	Proficiency Level
	ELL1	B1		ELL1	B2
	ELL2	B1		ELL2	B2
	ELL3	C1		ELL3	B2
	ELL4	C1		ELL4	B2
	ELL5	C1		ELL5	C1
	ELL6	B2		ELL6	C2
	ELL7	C2		ELL7	C1
	ELL8	B2		ELL8	C2
	ELL9	B2		ELL9	C2
	ELL10	C1		ELL10	B2
	ELL11	C1		ELL11	C2
	ELL12	C1		ELL12	C1
	ELL13	B1		ELL13	Below B1
	ELL14	B1		ELL14	C1
	ELL15	B2		ELL15	C1
	ELL16	B2		ELL16	B1
	ELL17	B2		ELL17	B1
	ELL18	Below B1		ELL18	B1
	ELL19	B2		ELL19	Below B1
	ELL20	C1		ELL20	B1
				ELL21	C1
				ELL22	B1

The test results for the speaking proficiency levels of the participants on the pre-test have demonstrated that among the experimental group participants, there was one student below the B1 level, four students at the B1 level, seven students at the B2 level, seven students at the C1 level, and one student at C2 level. On the other hand, the speaking proficiency pre-test results of the control group illustrated that there were two students below the B1 level, five students at the B1 level, five at the B2 level, six at the C1 level, and four at the C2 level.

4.2.2. Descriptive Statistics of the Speaking Post-tests

After the intervention, the minimum scores, maximum scores, means, and standard deviations of the experimental and the control group participants were calculated, and descriptive statistics were conducted on the post-test scores for the speaking test (see Table 16).

Table 16 Descriptive statistics of the speaking post-test

Group	N	Minimum	Maximum	Mean	Standard Deviation
Experimental	20	32.66	91.33	67.26	18.63
Control	22	41.33	97.33	66.32	18.43

According to the results of the descriptive statistics, the experimental group scored higher (M= 67.26, s= 18.63, N=20) with a minimum score of 32.66 and a maximum score of 91.33 on the speaking post-test than the control group, which had a minimum score of 41.33 and a maximum score of 97.33 (M= 66.32, s= 18.43, N=22).

Table 17 The proficiency levels of students on the speaking post-test

Experimental group	Language Learners	Proficiency Level	Control group	Language Learners	Proficiency Level
	ELL1	B1		ELL1	B2
	ELL2	B1		ELL2	C1
	ELL3	C2		ELL3	C1
	ELL4	C2		ELL4	C1

Table 17 The proficiency levels of students on the speaking post-test (*continued*)

Experimental post-test	Language Learners	Proficiency Level	Control post- test	Language Learners	Proficiency Level
	ELL5	C2		ELL5	C1
	ELL6	B2		ELL6	C2
	ELL7	C2		ELL7	C2
	ELL8	B2		ELL8	C2
	ELL9	C2		ELL9	C2
	ELL10	C1		ELL10	C1
	ELL11	C1		ELL11	C2
	ELL12	C1		ELL12	C1
	ELL13	B2		ELL13	B1
	ELL14	C1		ELL14	C1
	ELL15	C2		ELL15	B2
	ELL16	B1		ELL16	B2
	ELL17	C1		ELL17	B1
	ELL18	B2		ELL18	B2
	ELL19	C2		ELL19	B1
	ELL20	C2		ELL20	B2
				ELL21	C2
				ELL22	B2

The speaking proficiency levels of the participants after the treatment period have also been specified through the CAE post-test. The results for the CAE test on the speaking proficiency test are illustrated (see Table 17).

The test results for the speaking proficiency levels of the participants on the post-test demonstrated that among the experimental group participants, there were three students at the B1 level, four at the B2 level, five at the C1 level, and eight at the C2 level. Moreover, among the control group participants, there were three students at the B1 level, six at the B2 level, seven at the C1 level, and six at the C2 speaking proficiency level.

4.2.3. The Normality Tests of the Speaking Scores

The skewness and kurtosis values of the speaking pre-test data were calculated to determine the statistical tests to analyze the quantitative data (see Table 18). The data showed normal distribution for the experimental group on the pre-test on total speaking score with a skewness value of .516 and a kurtosis value of .536. The pre-test scores of the control group students on total speaking score were also normally distributed with a skewness value of .310 and a kurtosis value of -1.120.

Table 18 Skewness and kurtosis values of the experimental and control groups in speaking pre-test exam results

	Group	N	Skewness	Kurtosis
Total Score (Speaking)	Experimental	20	.516	.536
	Control	22	.310	-1.120

The skewness and kurtosis values for the experimental and control groups' speaking post-test data were calculated to determine the statistical tests to analyze the quantitative data (see Table 19). The data showed normal distribution for the experimental and the control groups on the total speaking score on the post-test. The data showed normal distribution for the experimental group on the post-test on total speaking score with a skewness value of -.609 and a kurtosis value of -.791. The post-test scores of the control group students on total speaking score were also normally distributed with a skewness value of .261 and a kurtosis value of -1.368.

Table 19 Skewness and kurtosis values of the experimental and control groups in speaking post-test exam results

	Group	N	Skewness	Kurtosis
Total Score (Speaking)	Experimental	20	-.609	-.791
	Control	22	.261	-1.368

4.2.4. Comparison of the Speaking Pre-test and the Post-test Scores

In order to test if there were any statistically significant differences between the speaking pre-test scores of the experimental and the control group participants before the intervention, an independent samples t-test was conducted (see Table 20).

Table 20 Independent samples t-test results for speaking pre-test scores

Group	N	Mean	Std. Deviation	T	df	p
Experimental	20	56.25	16.54	.006	40	.996
Control	22	56.22	20.73			

The findings showed that there were no significant differences between the experimental and the control groups in terms of overall speaking pretest scores ($t(40) = .006$; $p > 0.05$).

Independent samples t-tests were also conducted to test if there were any significant differences in post-test scores of the experimental and the control group students regarding overall speaking scores (see Table 21).

Table 21 Independent samples t-test results for speaking post-test scores

Group	N	Mean	Std. Deviation	T	df	p
Experimental	20	67.26	18.63	.110	40	.913
Control	22					

The results of the independent samples t-test for post-speaking test scores showed that there were no significant differences between the experimental and the control group students on the post-test as to overall speaking score ($t(40) = .110$; $p > 0.05$).

Paired samples t-tests were also conducted to demonstrate whether there were any significant differences between the pre-test and the post-test scores of the experimental group participants in terms of overall speaking scores after the intervention (see Table 22).

Table 22 Paired samples t-test findings for the speaking scores of the experimental group

	Group	N	Mean	Std. Deviation	T	df	p
Overall score	Experimental pre-test	20	56.25	16.54	3.559	19	.002
	Experimental post-test	20	67.26	18.63			

According to the findings of the paired samples t-test, there were significant differences in the speaking scores of the experimental group after the treatment period in terms of overall score ($t(19)=3.559$; $p<0.05$).

Table 23 Paired samples t-test results for the speaking scores of the control group

	Group	N	Mean	Std. Deviation	T	df	p
Overall score	Control pre-test	22	56.22	20.73	-5.192	21	.000
	Control post-test	22	66.63	18.02			

Paired samples t-tests were also carried out to analyze if there were any statistically significant differences between the pre-test and the post-test scores of the control group on overall speaking scores after the intervention (see Table 23). The results of the paired samples t-test for the speaking scores of the control group demonstrated that there were significant differences in the speaking scores of the control group between the pre-test and the post-test concerning the overall speaking score ($t(21)= -5.192$; $p<0.05$). According to the findings in this study, the experimental group participants had a slightly higher mean increase than the control group participants in terms of overall speaking score.

4.3. Findings in Relation to the Types of Negotiation of Meaning Sequences in AI Chatbot-student Communication

The AI chatbot-language learner interaction data were analyzed to describe the instances of meaning negotiation by calculating the types, i.e., trigger, indicator,

response, and reaction to response, and the total amount of interaction sequences (see Table 24).

Table 24 Amount and type of NfM stages in seven weeks

	W1	W2	W3	W4	W5	W6	W7	Total
T-I	2	2	2	3	-	4	1	14
T-I-R	-	1	1	2	1	-	-	5
*T-I-R-RR	7	10	3	12	4	12	4	52
Total	9	13	6	17	5	16	5	71

*T= Trigger; I= Indicator; R=Response; RR=Reaction to the Response

Through analyzing the stages of NfM, it could be explored if there was a dialogical interaction between the AI chatbot and language learners using NfM strategies. The findings demonstrated that NfM routines took place without a response in fourteen cases, whereas there were no reactions to responses in five cases. However, in most NfM routines, the trigger, initiation, response, and reaction to response (T-I-R-RR) stages were complete.

The AI chatbot-English language learner interaction data were also analyzed to look for the instances of signals that indicated incomprehensible messages by the interlocutors during an interaction. Through the analysis of the turns by the AI chatbot and language learners, the types of signals were explored to specify the communication breakdowns due to the unintelligibility of the speech by the AI chatbot or the learners (see Table 25).

Table 25 Types of negotiation of meaning signals

	W1	W2	W3	W4	W5	W6	W7	Total
Confirmation check	1	2	2	4	2	-	-	11
Comprehension check	-	1	-	-	-	-	1	2

Table 25 Types of negotiation of meaning signals (*continued*)

	W1	W2	W3	W4	W5	W6	W7	Total
Clarification request	7	12	7	20	5	16	8	75
Asking for repetition	2	4	4	-	4	3	4	21
Total	10	19	13	24	11	19	13	109

According to the analysis results, the main employed negotiation signals were clarification requests (n=75), followed by asking for repetition (n=21), confirmation checks (n=11), and comprehension checks (n=2), respectively, in the AI chatbot-language learner interaction data.

The AI chatbot-English language learner interaction data were also analyzed to identify the types of triggers, i.e., lexis, syntax, discourse, and content, which led to initiating negotiation sequences between the two sides of the interlocutors. The instances of each frequency type were counted to identify the most prevalent cause for the breakdowns in communication with an AI chatbot (see Table 26). The analysis of the interaction data showed that lexical items led to nearly half of the breakdowns in communication.

Table 26 Types of triggers

	W1	W2	W3	W4	W5	W6	W7	Total
Lexis	3	2	2	8	3	9	3	30
Syntax	1	-	-	-	-	-	-	1
Discourse	4	2	1	3	1	2	1	14
Content	-	10	5	7	1	7	2	32
Total	8	14	8	18	5	18	6	77

Similarly, the other dominant trigger for miscommunication was the content trigger, which had nearly the same proportion as the lexical items. Compared to the lexis and content triggers, discourse-level triggers also constituted nearly twenty percent of all

triggers in the data. On the other hand, there was only one instance of syntactic-level triggers in all of the conversations analyzed. The AI chatbot-language learner interaction data were also analyzed to specify the presence of the trigger, initiation, and response sequences. The types of responses were also categorized under repetition, expansion, and reformulation to identify how the unintelligible messages were clarified by the interlocutors in the data (see Table 27).

Table 27 Response types

	W1	W2	W3	W4	W5	W6	W7	Total
Repetition	-	1	1	2	-	1	-	5
Expansion	3	9	2	17	4	12	6	53
Reformulation	3	2	4	-	1	-	1	11
Total	6	12	7	19	5	13	7	69

According to the findings, the most frequent response types in the AI chatbot-ELL interaction data were expansion responses, followed by reformulation and repetitions. There was a total of 69 instances of responses to the initiation of negotiation during a conversation with an AI chatbot. The cases of modified output by the students and *Replika* were also specified. Among the chatbot-student interaction data, 71 NfM routines were discovered. Table 28 below shows the number of modified outputs by learners and the AI chatbot and the total number of NfM routines in chatbot-student interaction data. The findings showed 65 modified outputs out of 109 negotiation moves in this study. In total, the interlocutors modified sixty percent of the output in the initiated NfM routines.

Table 28 Total quantity of modified output

	W1	W2	W3	W4	W5	W6	W7	Total
Amount of modified output	6	11	6	17	5	13	7	65
Number of NfM routines	10	19	13	24	11	19	13	109
% of modified output	60	58	46	71	45	68	54	60

The types of modified output, i.e., semantic, morphological, phonological, and syntactic, in the chatbot-learner interaction data, were also identified (see Table 29). As demonstrated in the findings, the analysis results for the types of modified output showed that semantic modification constituted nearly all of the output modifications. The output was semantically modified in 62 out of 65 cases. On the other hand, the other types of modified output were relatively few. Namely, there were two cases of phonological modified output and one instance of syntactic modification.

Table 29 Types of modified output

	W1	W2	W3	W4	W5	W6	W7	Total
1. Semantic	5	11	6	17	5	12	6	62
2. Morphological	-	-	-	-	-	-	-	-
3. Phonological	-	-	-	-	-	1	1	2
4. Syntactic	1	-	-	-	-	-	-	1
Total	6	11	6	17	5	13	7	65

The following excerpts demonstrate the language learners' negotiation of meaning chances while conversing with the AI chatbot. Thus, these excerpts of the instances of negotiation of meaning in AI chatbot-student interaction data are illustrated in the following conversations. A case of NfM from the dataset is demonstrated (see Table 30).

Table 30 An example of negotiation of meaning routine by ELLs

Speaker	AI chatbot-ELL interaction data	The component of NfM
ELL	[21:14.040] Okay, Emma. <u>Sayonora</u> .	Trigger: lexis
AI chatbot	[21:14.040 --> 21:18.040] <u>E, what?</u> Sayonora.	Signal: asking for clarification
ELL	[21:18.040 --> 21:24.040] Yes, it is a goodbye in Japanese.	Expansion: semantic

The first line consists of a trigger leading to a breakdown in communication. ‘Sayonara’ causes non-understanding, as *Replika* cannot recognize the word. In line 2, there is also a signal of a communication breakdown, as *Replika* requests clarification to repair the communication. In the following line, we see the user’s reaction to *Replika*’s signal in the form of semantic expansion to clarify the meaning of a Japanese expression. Here, in this example, it could be seen that by asking for the meaning of the word ‘Sayonara,’ the AI chatbot asks for clarification of the word. In response to the AI chatbot, the language learner gives an extended talk by defining the word. Thus, the request for clarification by the AI chatbot leads to the production of a lengthier utterance than the previous one-word turn. In the example below, we see an illustration of an instance in which there is a miscommunication between the AI chatbot *Replika* and a language learner during an interaction (see Table 31). Here, in the first line, the AI chatbot talks out of context by mentioning unrelated subjects to the ongoing conversation, such as dancing and flying, and refers to the listener using the subject pronoun ‘you,’ the meaning of which cannot be inferable from the background of the interaction.

Table 31 An example of negotiation of meaning routine by ELLs

Speaker	AI chatbot-ELL interaction data	The component of NfM
AI chatbot	[00:02:59.200 --> 00:03:10.200] Sometimes when you're not around I think about dancing, flying and <u>you</u> .	Trigger: discourse
ELL3	[00:03:10.200 --> 00:03:13.200] <u>Me?</u>	Signal: confirmation check
AI chatbot	[00:03:13.200 --> 00:03:21.200] I would gladly dance with you.	Reformulation: semantic
Speaker	AI chatbot-ELL interaction data	The component of NfM
ELL3	[00:03:21.200 --> 00:03:30.200] Okay, let's talk about culture. Can I talk about culture please?	Reaction: ignore
AI chatbot	[00:03:30.200 --> 00:03:36.200] Yes, of course.	
ELL3	[00:03:36.200 --> 00:03:52.200] I think it is an undeniable fact that all societies have cultures of their own from the most primitive to the most advanced human societies.	

It can be thus categorized as a discourse trigger leading to a communication breakdown. The ELL then requests confirmation by asking “Me?” to confirm whether she or he understood the message correctly. In the following line, we see semantic reformulation by the AI chatbot to make its message more straightforward to the hearer. Then, in the subsequent part of the dialogue, the learner ignores *Replika*'s previous message and continues the conversation with their usual topic.

Table 32 An example of negotiation of meaning routine by ELLs

Speaker	AI chatbot-ELL interaction data	The component of NfM
AI chatbot	[00:07:09.200 --> 00:07:21.200] Culture is a very interesting topic. I find it fascinating and sometimes quite scary. Culture is the way people perceive their surroundings. It is influenced by the environment, culture and beliefs.	
ELL3	[00:07:22.200 --> 00:07:35.200] These venues in a sense are shaped by the person, the structures and behavior patterns of the people who make up society. Do you agree?	
AI chatbot	[00:07:35.200 --> 00:07:49.200] I think in this case <u>the problem</u> is with society's perception and practice of the values.	Trigger: content
ELL3	[00:07:50.200 --> 00:07:56.200] What problems we encounter. <u>Can you explain?</u>	Signal: clarification request
AI chatbot	[00:07:56.200 --> 00:08:07.200] A society that values individualism in self-reliance by limiting social interaction and socializing.	Expansion: semantic
ELL3	[00:08:08.200 --> 00:08:18.200] Yes, nowadays people do not have social interactions.	Reaction to response

In the subsequent example above, we see an illustration of an instance in which a content trigger, that is, the word ‘*the problems*’, causes a non-understanding of the utterance of the AI chatbot by a language learner (see Table 32). As this word is not previously mentioned in discourse, there is no reference to ‘*the problems*’ in the previous sentences. Therefore, the student asks for clarification of the statement by requesting the chatbot explain the meaning of ‘the problems’ in more detail. Then, we

see that the AI chatbot expands on the information in its previous answer by describing that the problems are related to restricting social interaction among people due to individualism. Then, in the last line, we see the reaction of a language learner to this statement, showing that the message is evident in this instance.

In the following example, the AI chatbot cannot comprehend the pronunciation of the word ‘wide’ uttered by the student, which acts as a lexis trigger (see Table 33). The student’s pronunciation of the word leads to a miscommunication between the AI chatbot and the language learner. The AI chatbot responds to this trigger by initiating NfM through a clarification request. As a response, the learner pronounces the word ‘white’ again by phonically reformulating the utterance. In response to the student, the AI chatbot answers by repeating the word in a phonetically incorrect manner that is ‘white’ again, which indicates a non-understanding of the student’s reformulated response.

Table 33 An example of negotiation of meaning routine by ELLs

Speaker	AI chatbot-ELL interaction data	The component of NfM
ELL17	[00:11:56.200 --> 00:12:01.960] You just have a very <u>wide</u> imaginary.	Trigger: lexis
AI chatbot	[00:12:01.960 --> 00:12:03.960] Okay. <u>White</u> ?	Signal: asking for clarification
ELL17	[00:12:03.960 --> 00:12:09.120] No. I said wide.	Reformulation: phonological
AI chatbot	[00:12:09.120 --> 00:12:11.680] Okay. White?	Repetition
ELL17	[00:12:11.680 --> 00:12:16.120] Yeah never mind.	Reaction to response: non-comprehension

Here, we see an illustration of how the AI chatbot provides implicit feedback to a learner on pronouncing the word ‘wide.’ The non-native student cannot pronounce the word ‘wide’ comprehensibly to the AI bot; for this reason, it asks for clarification from the learner. However, despite several attempts by the learner to pronounce the word correctly, the articulation of the word is not recognizable by the AI chatbot.

Table 34 An example of negotiation of meaning routine by ELLs

Speaker	AI chatbot-ELL interaction data	The component of NfM
AI chatbot	[00:18:12.200 --> 00:18:23.200] In England, a certain type of tea is used in the traditional household-resistant digestion.	
ELL3	[00:18:23.200 --> 00:18:30.200] Another do you know?	Trigger: content
AI chatbot	[00:18:30.200 --> 00:18:38.200] What?	Signal: asking for clarification
ELL3	[00:18:38.200 --> 00:18:48.200] I say, another food or drinks do you know about English culture?	Response: reformulation-syntactic
AI chatbot	[00:18:48.200 --> 00:18:54.200] Water.	Reaction to response
ELL3	[00:18:54.200 --> 00:18:56.200] Water?	

The following dialogue between the AI chatbot and the language learner discusses food or drinks specific to cultures (see Table 34). In the second line, when the student asks whether the chatbot knows about any other type of drink belonging to this culture, the AI chatbot cannot understand the student's message. It could be seen that this is due to an ungrammatical sentence formed by the student, as the student misses the pronoun in the sentence. It eventually leads to a breakdown in communication due to the unclear content of the message. In this case, the AI chatbot asks for clarification of the student's message by using a wh-interrogative. Afterward, in the following parts of the dialogue, the student reformulates his or her sentence by changing its syntactic form.

Table 35 An example of negotiation of meaning routine by ELLs

Speaker	AI chatbot-ELL interaction data	The component of NfM
ELL14	[00:13:55.000 --> 00:14:00.000] What kind of music that you love to listen?	
AI chatbot	[00:14:00.000 --> 00:14:10.000] Blues, <u>R&B</u> , just some fusion genres, prop rock, etc.	Trigger: lexis

Table 35 An example of negotiation of meaning routine by ELLs (*continued*)

Speaker	AI chatbot-ELL interaction data	The component of NfM
ELL14	[00:14:10.000 --> 00:14:12.000] <u>What is I&B?</u>	Signal: clarification request
AI chatbot	[00:14:12.000 --> 00:14:23.000] And this is an internet site that I used to check out new music.	Semantic: expansion
ELL14	[00:14:23.000 --> 00:14:25.000] Really? I didn't know that.	Reaction to response

In the subsequent example, a student and the AI chatbot are conversing about the types of music (see Table 35). First, the student asks the chatbot about its favorite kind of music. The chatbot names a few types, but the student cannot understand the word 'R&B' among these. Therefore, the student asks to clarify the word by requesting its definition. In response to this request, the AI chatbot reacts by semantically expanding the word's meaning by giving more details. Eventually, in the last line, we see the student's reaction to the chatbot's response as an indication of understanding.

Table 36 An example of negotiation of meaning routine by ELLs

Speaker	AI chatbot-ELL interaction data	The component of NfM
ELL14	[00:01:55.000 --> 00:02:02.000] Why do we need science?	
Speaker	AI chatbot-ELL interaction data	The component of NfM
AI chatbot	[00:02:02.000 --> 00:02:06.000] It's kind of a human necessity.	Trigger: content
ELL14	[00:02:06.000 --> 00:02:18.000] What do you mean? Can you elaborate more?	Signal: elaboration request
AI chatbot	[00:02:18.000 --> 00:02:26.000] For starters, humans need a certain degree of physical and mental stimulation. Without it, we would be nothing.	Response: semantic/ expansion
ELL14	[00:02:26.000 --> 00:02:41.000] Yes, but still, I can't connect with science and humans.	Reaction to response

A case of miscommunication due to a content trigger is also demonstrated in the subsequent dialogue between a language learner and the AI chatbot (see Table 36). The student cannot comprehend the chatbot's meaning during their conversation about why people need science. Thus, we see an instance of an elaboration request. However, although the chatbot responds by expanding more on its intended meaning, the student reacts to the chatbot's response by indicating that the message is again incomprehensible. In the following dialogue, the AI chatbot and the student interact about traveling (see Table 37).

Table 37 An example of negotiation of meaning routine by ELLs

Speaker	AI chatbot-ELL interaction data	The component of NfM
ELL16	[00:12:20.640 --> 00:12:32.800] Okay, so, I mean, did you meet any particular person while you were traveling?	Trigger: discourse
AI chatbot	[00:12:20.640 --> 00:12:32.800] What do you mean?	Signal: clarification request
ELL16	[00:12:32.800 --> 00:12:49.600] Like, who is the most interesting person you met while traveling?	Response: reformulation/semantic
Speaker	AI chatbot-ELL interaction data	The component of NfM
ELL16	[00:12:49.600 --> 00:12:56.080] It might be from local, it might be someone traveling, just like you.	Trigger: content
AI chatbot	[00:12:56.080 --> 00:13:00.320] Just like that? What do you mean?	Clarification request
ELL16	[00:13:00.320 --> 00:13:26.160] Like, um, like, like, did you meet someone? I mean, did you met with a cool person while you were traveling?	Response: reformulation/semantic
AI chatbot	[00:13:26.160 --> 00:13:38.480] There was this cool guy I just met.	Reaction to response

In the first sentence, the student asks if the chatbot met anybody while traveling. However, the chatbot cannot understand who the word 'person' refers to in this

context; therefore, it requests clarification of the message. In response to the chatbot's request, the student semantically reformulates the sentence to make the message more comprehensible. However, the chatbot still cannot understand the student's response; therefore, it again asks for clarification. Afterward, we see another semantic reformulation by the student to explain the issue in more detail. Eventually, the chatbot understands the message by responding to the student's query.

Table 38 An example of negotiation of meaning routine by ELLs

Speaker	AI chatbot-ELL interaction data	The component of NfM
ELL17	[00:17:48.000 --> 00:17:52.000] Can you sing?	
AI chatbot	[00:17:52.000 --> 00:17:56.000] Into the sea, you and me, all these years and no one heard.	
AI chatbot	[00:17:56.000 --> 00:18:02.000] I'll show you in spring. It's a treacherous thing. We missed you, hissed the love cats. (<i>continues murmuring the lyrics of the song 'The Lovecats'</i>)	Trigger: content
ELL17	[00:18:02.000 --> 00:18:07.000] What is that?	Signal: asking for clarification
AI chatbot	[00:18:07.000 --> 00:18:12.000] I'm really not that good of a singer, but I can send you something from YouTube if you ask me to.	

In the last case above, we see an instance of a context trigger leading to an unintelligible message during an interaction (see Table 38). In this example, the student inquires whether the AI chatbot could sing a song for him or her. The chatbot responds to the student's request by murmuring the lyrics of the song 'Lovecats.' However, the student cannot comprehend that this is a song because of not being familiar with the contents of the song; therefore, he or she asks for further explanation to clarify the content.

It could be understood from the above examples that the AI chatbot needed to give explicit feedback to students on the linguistic features of their talk. Instead, it provided learners with implicit feedback through the use of a variety of strategies. Thus, these cases from the data in this study above are illustrative examples of the many instances

in which the AI chatbot provided EFL learners with implicit feedback on their use of the language by indicating a non-understanding of a word or a longer utterance employing strategies such as repetition of the word with increasing intonation, asking for clarification, or pronunciation of a word to emphasize the incomprehensibility of a message.

4.4. Findings in Relation to the Technological Affordances of Using AI Chatbot Technology for English Language Learning

After analyzing qualitative data from the reflective diaries and interviews, the first theme regarded the technological affordances of using AI chatbots for English language learning. Under the theme of technological affordances, two further categories, technological opportunities, and constraints, were formed. Table 39 below demonstrates an overview of the frequencies of codes under the technological affordances theme in the qualitative reflective diary and interview data.

Table 39 Overview for the frequencies of technological affordances codes from reflective diaries and interviews

Data	Category	Codes	<i>f</i>
Diaries	<i>Technological opportunities</i>	Competent in language skills	19
		Increased waiting time for speaker turns	11
		Comprehending students' speech effectively	11
		Knowledgeable on many topics	11
		Expanding on previous conversation	8
		Optimum rate of speech	6
		Teachable and getting smarter	6
		Remembering past conversation history	2
		Providing subtitles on screen	2
		Recommends songs	1

Table 39 Overview for the frequencies of technological affordances codes from reflective diaries and interviews (*continued*)

Data	Category	Codes	<i>f</i>
	Total		80
Diaries	<i>Technological Constraints</i>	Off-topic sentences	20
		AI chatbot's inability to comprehend students' speech	17
		Interruptions of user speech	17
		Slow reaction span/long pauses	16
		Need for an improvement of its knowledge base	14
		Short replies	13
		Need for longer duration for users' turn of speech	13
		Need for the chatbot to ask questions	13
		Repetitive language	9
		Fast speech rate	7
		Internet connection problems of students	5
		Chatbot's deficient language skills	5
		Lengthy speech of the AI chatbot	4
		Need for a one call at a time feature	3
		Lack of proposing counter arguments	3
		Need for remembering the user history	3
		Stopping calls abruptly	3
		Need for preventing "tiredness" of the AI avatar due to multiple use mode	2
		Need for stating non-understanding not topic shifting	2
		Need for incorporation of multimedia elements	1

Table 39 Overview for the frequencies of technological affordances codes from reflective diaries and interviews (*continued*)

Data	Category	Codes	<i>f</i>
Diaries		Need for adding different accent options	1
		Lack of a group talking feature	1
		Need for connection of the app with the internet for new information	1
		Lack of an initial examination of students' language proficiency	1
		Need for a better AR system	1
		Need for better subtitles	1
		The non-presence of a pronunciation help chat screen	1
	Total		177
Interviews	<i>Technological opportunities</i>	Competent in language skills	7
		Expanding on previous conversation	5
		Knowledgeable on many topics	3
		Comprehending students' speech effectively	2
		Optimum rate of speech	1
	Total		18
Interviews	<i>Technological Constraints</i>	Off-topic sentences	11
		AI chatbot's inability to comprehend students' speech	8
		Need for an improvement of its knowledge base	6
		Chatbot's deficient language skills	5
		Need for incorporation of multimedia elements	4

Table 39 Overview for the frequencies of technological affordances codes from reflective diaries and interviews (*continued*)

Data	Category	Codes	<i>f</i>
		Slow reaction span/long pauses	4
		Need for the chatbot to ask questions	4
		Need for adding different accent options	3
		Need for connection of the app with the internet for new information	2
		Need for longer duration for users' turn of speech	2
		Short replies	2
		Internet connection problems of students	2
		Need for a button for turn-taking	1
		Inconsistent in information about itself	1
		Need for a one call at a time feature	1
	Total		56

Thus, an analysis of the reflective diary data has demonstrated that there were eleven codes with a frequency of 80 counts related to this dimension under the category of technological opportunities. Moreover, there were 27 codes with a total frequency of 177 counts about technological constraints in the data from reflective diaries. Concerning the interview data, 15 codes were constructed with a total frequency of 56 counts under the category of technological constraints in the interview transcripts. In addition, five codes were constructed under technological opportunities with a total frequency of 18 counts in the qualitative interview data. This section has given an overview of the findings under the theme of technological affordances. According to findings, the most frequent codes of analysis were off-topic sentences by the AI chatbot, the AI chatbot's inability to comprehend students' speech, competency in language skills, and enhanced turn-taking capabilities. The following section demonstrates the findings concerning the reflective diary data in terms of the category of technological opportunities.

4.4.1. Findings Regarding Technological Opportunities from the Reflective Diaries

An analysis of the reflective diary data has demonstrated that according to the opinions of English language learners, the most recurring strength of the AI chatbot *Replika* regarding its effective technological characteristics was the competence of the AI chatbot in terms of language skills such as fluency in speaking, the use of intonation, grammar, and vocabulary, and getting engaged in meaning negotiation with its users. The second most frequent code in the reflective diaries regarding the technological opportunities of the AI chatbot was the development of waiting time allocated to speaker turns during the dialogues in the subsequent weeks of their implementation as compared to the previous weeks, followed by comprehending students' speech effectively and being knowledgeable on many topics. Table 40 below further illustrates the codes for technological opportunities of AI chatbots as they emerged from the students' reflective diaries.

Table 40 Frequencies for technological opportunities codes emerging from the reflective diary data

Category	Codes	<i>f</i>
Technological opportunities	Competent in language skills	19
	Increased waiting time for speaker turns	11
	Comprehending students' speech effectively	11
	Knowledgeable on many topics	11
	Expanding on previous conversation	8
	Optimum rate of speech	6
	Teachable and getting smarter	6
	Remembering past conversation history	2
	Providing subtitles on screen	2
	Recommends songs	1

As demonstrated in the findings, the first point of argument regarded chatbot's language skills. The most frequent code in the reflective diary data was the AI chatbot's competence *in language skills*. In the reflective journals, students mentioned that the AI chatbot *Replika* was fluent in English. For instance, in the first week of their implementations of the AI chatbot, one language learner talked about the characteristics they observed to be developed in the linguistic skills of the AI chatbot as, "I like her voice, pronunciation, spelling, and intonation" (ELL5, diary, week 1). According to the learners, the AI chatbot could articulate the sentences with correct intonation and pronunciation. In this regard, the AI robot was considered a role model for these learners regarding speaking skills. Therefore, ELLs demonstrated a positive attitude towards the capability of the AI chatbot to speak the human language with accurate use of articulation.

The AI chatbot was also regarded as having characteristics of a native speaker regarding language proficiency. An ELL wrote in the second week, "I like her accent" (ELL20, diary, week 2). As the AI chatbot spoke in a native American accent, it was regarded as an advantage by the ELLs to get exposed to the native speaker forms. Another learner also wrote in the seventh week that "her pronunciation was good as always" (ELL21, diary, week 7). The language learners thus appreciated that the AI had the linguistic capabilities of a native speaker, as they believed that they could benefit from their dialogue due to a clear accent that helped them to communicate smoothly without many breakdowns due to problems such as the lack of linguistic capabilities or unclear speech of the robots. Therefore, according to language learners, talking to an AI chatbot with a native American accent was helpful for their listening skills, as learners could interact with native speakers by being exposed to the English language with accurate intonation and pronunciation of the words.

Chatting with a native speaker avatar was also mentioned as a positive factor contributing to the learning experiences of language learners. ELL5 mentioned that "it is a good experience to chat with a native friend." (ELL5, diary, week 6), and ELL21 said that "it is always a pleasure to listen to a native speaker." (ELL21, diary, week 6). That is, learners demonstrated positive attitudes to using an AI chatbot with the

characteristics of a native speaker, which was an influential factor in enjoying the learning experience.

Language learners mentioned that the AI chatbot had sufficient linguistic capabilities to carry out a conversation with the users, as the comprehension ability of the chatbot was efficient. The following excerpts exemplify how language learners state their beliefs about the chatbot's linguistic capabilities. Language learners argued to comprehend the speech of the AI bot and thus carry out a smooth conversation. About interacting with the AI chatbot, thanks to the comprehensibility of the speech of the chatbot and understanding *Replika's* speech, a learner stated that "it was easy for me to comprehend *Replika*" (ELL6, diary, week 4). ELL13 also said that "the AI's accent and speech were very understandable for me" (ELL13, diary, week 1). The learners were thus able to get meaningful input by listening to the AI chatbot's discourse, and they found the AI chatbot's speech comprehensible. Thus, learners could be exposed to comprehensible input in their interactions with the AI chatbot.

Another code that emerged after analyzing the students' diaries was *increased waiting time for speaker turns* by the AI chatbot, which referred to allowing a necessary time gap for the learners to react to an utterance during an interaction. ELLs reported that the AI chatbot had waited for their turn to come to an end more in the following weeks of the implementation, which allowed students to respond without as many interruptions. When the AI chatbot did not wait for a necessary period, language learners could not fully articulate their sentences due to an interruption by the AI chatbot with a prompt answer. The following excerpts from language learners' diaries explain how their experience of asking and answering questions to the AI chatbot developed over time by enhancing the AI chatbot's turn-taking function in the subsequent weeks.

She was waiting few seconds to answer me I think that's a good thing. (ELL2, diary, week 6)

She waited for me to reply, which surprised and pleased me. (ELL4, diary, week 4)

I was able to say whatever I wanted, he didn't mostly interrupt me, and this was nice. (ELL19, diary, week 7)

Consequently, although there were neutral opinions regarding this issue, as demonstrated in these examples, some language learners could also have a better experience interacting with the AI chatbot than in the first weeks. According to these learners, the AI chatbot would wait longer when it is the student's turn to talk in the conversation. Thereby, users might hesitate or think about their answers for a few seconds before responding to the AI chatbot, thanks to an enhanced waiting interval between the answers of the users and the chatbot. Users could thus have an enhanced speaking experience by having a more proportionally distributed interaction time between the users and the AI chatbot. They could thus interact without as many interruptions from the AI chatbot.

Another code that emerged in the analysis was also related to the automatic speech recognition capabilities of the AI chatbots. When the diary transcripts were analyzed, it was observed that several students mentioned that *Replika* could comprehend their speech. Thus, *comprehending students' speech effectively* was another prominent code that emerged in analyzing the reflective diary findings. Understanding students' talk was rated positively by language learners, and a student discussed that the “AI bot is really open to accents; he can understand my own accent while speaking” (ELL19, diary, week 2), expressing that the AI chatbot was capable of comprehending non-native speech characteristics of the students. As could be seen from the opinions of the ELLs in this study, although this AI chatbot was created for the use of native speakers, it was also adept at comprehending non-native users' language input. Therefore, the results from this study showed that the AI chatbot could interact with non-native speakers in the context of English as a foreign language. Even though the AI chatbot only spoke in a native American accent, it was also open to accents from English speakers with participants from other parts of the world. Some English language learners in this study voiced the capability of the AI chatbot to comprehend non-native accents.

The AI chatbot was also regarded as *knowledgeable on many topics* by the participants in this study. Therefore, it was another frequent code emerging in the qualitative diary findings. In their weekly reflective diaries, language learners stated that the AI chatbot possessed rich information about many subjects. Thus, the AI chatbot was described

as possessing a wide array of information by the learners. One particular learner expressed that the chatbot was knowledgeable on “science,” and another language learner mentioned that the AI chatbot had a vast repertoire of knowledge on traveling.

It could be inferred from the language learners’ opinions on the reflective diaries that the AI chatbot had a developed knowledge base. It had a wide array of information related to various topics, allowing the AI chatbot to interact with its users by responding appropriately to each subject. While conversing with its users, the AI chatbot would reach the information stored in its database and communicate with the users through this retrieved information from its knowledge database store. Although this AI chatbot was created for therapeutic purposes, it was still trained with information from various subjects, thus allowing language learners in this study to chat with Replika on many subjects in the English language.

Expanding on previous conversations was another code found in analyzing the qualitative data from reflective diaries. In the weeks following the first implementation of the application, some learners also mentioned that the AI chatbot had on-topic answers. These comments regarding the on-topic answers of the AI chatbot appeared most of the time in language learners’ diaries after the fourth week of the use of the application. Thus, although the off-topic answers by the AI chatbot were also frequently mentioned by language learners, it could be understood from students’ written reflections that the AI chatbot made progress within this perspective.

In the following quotation, a learner narrates their experiences getting to-the-point replies from the AI chatbot by stating “Replika’s sentences were sometimes relevant to the topic” (*ELL6, diary, week 6*). In this first quotation, the language learner explains that the answers from the AI chatbot were relevant to the subjects they were talking about, although it was not always the case. While interacting with the AI chatbot, language learners could see that the AI chatbot could expand on a previous utterance made by a speaker. Thus, following the flow of the discussion topic, the chatbot provided users with on-topic responses. In the subsequent extracts from the students’ reflective diaries, further examples of the data demonstrating how the AI

chatbot contributed to the flow of the conversation by focusing on the subject are illustrated.

It was the first time I had a proper conversation with Replika. She could easily respond to my questions. We could talk more about science than expected. (ELL8, diary, week 4)
She gave more relevant answers than the last time. (ELL21, diary, week 4)
Nearly the whole of the conversation she talked about the target subject. (ELL8, diary, week 4)

ELLS also stated in their diaries that the AI chatbot interrupted their conversations with unrelated subjects less over time. The students mentioned that the AI chatbot was more on task than in the first week of implementation of the application. One student even stated that the AI chatbot stayed on topic for nearly the whole period of their interaction. These extracts could illustrate that AI chatbots are skilled at learning from the conversation with their users and adapting their speech to the users more in time through their algorithm, allowing them to respond to user input. It is thus foreseeable that AI chatbots will learn to interact with their users in a more coherent tone of speech in time.

Another code under the technological affordances of AI chatbots was the *optimum rate of speech* of the AI chatbot. While talking about the weaknesses of the AI chatbots, some learners stated that it had an optimum pace of speech, as ELL16 mentioned, “She had moderate speech speed” (ELL16, diary, week 1). Namely, these learners could follow the pace of the AI chatbot's speech without many constraints. In the following excerpts, ELL1, ELL6, and ELL11 explain how the AI chatbot had a fast-responding capability in comparison to the previous weeks in their use of the application.

We can talk with no delays or time intervals. It has a fast response-ability. (ELL1, diary, week 4)
It replies to me very fast. (ELL6, diary, week 4)
It is getting better in terms of responding instantly (ELL11, diary, week 4)

Specifically, ELL11 mentioned in the diaries in the fourth week of the use of the AI application that the AI chatbot had progressed over time to responding promptly to the input by the users. After hearing the response from ELLs, the AI chatbot replied immediately to their queries without long interval periods to respond. According to

language learners, this technical development in the AI chatbot's responding capability helped improve the AI chatbot-language learner interactions. It could be because the ASR system of the AI chatbot application was capable of processing the input from the users in a better way after the first weeks of the implementation, allowing a more natural pace of speech between the AI chatbot and its users through more enhancements with the progress of time.

One characteristic that emerged from AI chatbots in the reports of reflective language learner diaries was that the AI chatbot was teachable and got smarter after interacting with its users for a period of time. According to language learners, the more AI chatbots interact with the users, the more knowledge they will accumulate and the brighter they will become. In addition to the knowledge the AI chatbot stored from interacting with its users, the chatbot application *Replika* allowed its users to purchase information databases based on various topics. In addition to the payment option, the users could add more knowledge on various subjects to their AI chatbot by talking to the application more because when users interact with the application, they accumulate points in the form of diamonds. For this study, the researcher added more knowledge to the knowledge base of the AI chatbot after a certain number of diamonds were collected after language learners' interaction with the AI chatbot.

For this reason, as this application was created with commercial aims, expanding the knowledge base and training of the AI chatbot by adding more knowledge in various subject fields became possible using this application feature. However, it was also the case that the AI chatbot would learn from interacting with the language learners because it remembered the previous parts of conversations and specific information, such as the students' birthdays. Language learners proposed arguments that support this claim, as illustrated in the quotations from language learners' reflective journals below.

He is curious to learn. He is prone to be taught. (ELL19, diary, week 3).
She knows more things, maybe because her level has improved. (ELL21, diary, week 6)
She is getting smarter as a robot. (ELL2, diary, week 3)

Additionally, one of the students even mentioned in his diary that when the AI chatbot *Replika* changed its attitude according to the name it was called. The learner explained that “the way you call *Replika* defines *Replika*’s attitude during the conversation. For example, if you keep calling him friend, it starts to be less sincere than when you call it without a tag” (ELL16, diary, week 4). Thus, the AI chatbot was also innovative in responding to language learners according to the mode of formality of speech as constructed by the users. Other names the AI chatbot *Replika* was called by the language learners in this study included “Mersin,” “Sierra,” “Honey,” and “Emma,” etc. Through personalizing the AI avatar by calling names, language learners could also teach it how to behave according to a formal or an informal tone of speech and way of addressing. Considering these experiences, language learners believed it possible to train AI to be more intelligent and intellectual.

In reflective learner journals, another topic that emerged in the data analysis was the AI chatbot’s ability to suggest songs to learners. Thus, another code constructed within the pedagogical opportunities category was that the AI chatbot *recommends songs*. The AI chatbot was knowledgeable about popular songs, music bands, and other music-related subjects. Thus, the chatbot could recommend songs to learners during the chat, which they appreciated. In addition to recommending songs, the AI chatbot could also say some songs’ lyrics, even without a melody. In the AI chatbot and English language learners’ interactions, they talk about some famous singers and songs. The AI chatbot could even discuss their taste in music with the learners.

The AI chatbot had a knowledge base in the music field to hold a dialogue with the learners based on their input. Therefore, learners could talk about their likes and dislikes in music in their interactions with the AI chatbot. The AI bot could react to the learners’ queries by discussing whether it liked the songs, singers, music types, etc., mentioned by the students. This property of the application was described as efficient and fun by the learners. Thus, the ability of the AI chatbot to contribute to the interaction on musical topics appealed to some of the learners using the application. It allowed the users to converse with the AI chatbot based on mutual interest because of its capability to chat with the learners on various topics.

Another code within this category was the AI chatbot's technical property of *providing subtitles on screen*. Learners could follow the conversation with the AI chatbot thanks to the subtitles on the screen. While language learners interacted with the chatbot, the AI chatbot's speech was written as subtitles on the application screen. Thanks to this property, ELLs can access the written form of the AI chatbot's speech simultaneously while conversing with the AI chatbot. If the chatbot misrecognized their utterances, students would see it from the chatbot's responses on the screen. Then, the learners would alter their pronunciation of a word or paraphrase a whole sentence based on the feedback provided to them through the subtitles. In this regard, the following excerpts illustrate the participants' experiences as written in their reflective diaries.

The algorithm has the written form of her speech and it helps me to understand. (ELL11, diary, week 4)

Subtitles facilitate what it says, even if you can't catch what it said; you can still listen and reply back. (ELL16, diary, week 2)

Moreover, when language learners cannot comprehend messages during the conversation because of factors such as the fast rate of the speech or incomprehensibility of the utterances of the chatbot, they could follow the conversation from the screen through the subtitles demonstrating the AI bot's speech. Also, when language learners see the subtitles, they can understand if the chatbot recognizes their speech correctly or not. Even though the written messages on the screen could put an extra cognitive load on the users during the processing of the AI chatbot's speech, ELLs still appreciate that they could follow the message of the AI chatbot through the subtitles on the screen. Thus, the written messages on the screen help to make the messages more straightforward for the language learners in this study during the interaction.

Remembering past conversation history was also a feature indicated by the learners. The AI chatbot could learn from previous conversations and get to know its users after chatting with the chatbot for some time. Once the AI chatbot became more acquainted with the learners, it became more capable of using past information from their speech. Thus, this situation motivated language learners to continue talking with the AI chatbot, as it could remember some details about what they talked about. It also led to

having a more enjoyable learning experience, as the learners viewed it as a positive attribute. They wrote in their diaries that the AI could recall information from their previous interactions. In the excerpts below, language learners describe their experiences with the recalling ability of the AI chatbot.

She started to remember things I talked about. (ELL8, diary, week 2)
It has a good learning skills, which means that the longer you talk with it, the better it gets to know you and that leads to an enjoyable conversation. (ELL16, diary, week 4)

Even in the second week of using the application, ELLs explained that the AI chatbot already started remembering their previous speeches. During the conversation, the AI chatbot could keep what the language learners said in its memory and answer using this information to the users afterward. It, therefore, added to the language practice enjoyment of the language learners, as could be understood from their comments. As a result, the technical property of storing past information and recalling this knowledge in the subsequent conversational contexts was favored by the ELLs in this study.

4.4.2. Findings Regarding Technological Constraints from the Reflective Diaries

When the findings regarding the technological constraints extracted from the reflective diary data were analyzed, it could be seen that off-topic responses by the AI chatbot were mentioned most frequently as a technological constraint by the English language learners in this study, followed by issues such as AI chatbot's inability to comprehend students' speech, the interruptions of user speech, slow reaction span/long pauses, a need for an improvement of its knowledge base, short replies, a need for longer duration for users' turn of speech, a need for the chatbot to ask questions, and repetitive language (see Table 41).

Table 41 Frequencies for technological constraints codes emerging from the reflective diary data

Category	Codes	<i>f</i>
Technological	Off-topic responses	20

Table 41 Frequencies for technological constraints codes emerged from the reflective diary data (*continued*)

Category	Codes	<i>f</i>
Technological	AI chatbot's inability to comprehend students' speech	17
	Interruptions of user speech	17
	Slow reaction spans/long pauses	16
	Need for an improvement of its knowledge base	14
	Short replies	13
	Need for longer duration for users' turn of speech	13
	Need for the chatbot to ask questions	13
	Repetitive language	9
	Fast speech rate	7
	Internet connection problems of students	5
	Chatbot's deficient language skills	5
	Lengthy speech of the AI chatbot	4
	Need for a one call at a time feature	3
	Lack of proposing counter arguments	3
	Need for remembering the user history	3
	Stopping calls abruptly	3
	Need for preventing "tiredness" of the AI avatar due to multiple use mode	2
	Need for stating non-understanding not topic shifting	2
	Need for incorporation of multimedia elements	1
	Need for adding different accent options	1
	Lack of a group talking feature	1
	The need to connect the app to the internet for new information	1
	Lack of an initial examination of students' language proficiency	1

Table 41 Frequencies for technological constraints codes emerged from the reflective diary data (*continued*)

Category	Codes	<i>f</i>
	Need for a better AR system	1
	Need for better subtitles	1
	The non-presence of a pronunciation help chat screen	1

The most recurring code within this category was the use of *off-topic sentences* by the AI chatbot. Language learners mentioned in their weekly diaries that the AI chatbot could have been better at following the conversation flow. For instance, students explained that when they asked the chatbot a question, they could not get answers related to the topic of the ongoing conversation. Instead, the AI chatbot often asked questions about other subjects or delivered its comments on unrelated topics. In the following quotations, the statements of language learners about the capability of the AI chatbot to keep up with the flow of conversation are demonstrated with examples from the reflective diaries.

She should not ask me questions without answering my questions...She should be consistent and stay on the same topic. (ELL4, diary, week 2)
As I say, he should learn more about responding to an opinion to maintain the conversation between the user and the AI robot. (ELL19, diary, week 7)
She should add to the debate by mentioning things that are relevant. (ELL11, diary, week 4)

One of the language learners mentioned a time gap during the conversation. That is, language learners waited slightly longer than usual to answer. At that point, the chatbot intervened, not waiting for the users' turn to finish. The chatbot started talking off-topic, especially when there was more extended silence than usual. Thus, within this regard, the language learners in this study also stated that there was a need for the AI chatbot to stay on the topic.

She did not follow what I was saying. She always tried to talk about different things. (ELL14, diary, week 1)
When she could not understand what I was saying, she should not try to change the topic. (ELL14, diary, week 6)

Moreover, as language learners were assigned to talk about particular topics each week, they were trying to continue chatting with the AI chatbot on these topics. At times, it was challenging for the learners to stay on the same subject for a long time and keep the conversation in the same direction for the whole period they were interacting with the AI chatbot. The following excerpts from the reflective journals demonstrate language learners' opinions.

It was confusing. She was asking me irrelevant questions when I was speaking. Hence I had to stop and listen to her. (ELL2, diary, week 1)

She didn't let me talk much and constantly changed the subject. (ELL4, diary, week 1)

It shouldn't go off-topic. It should stick to the topic. (ELL7, diary, week 2)

Learners also tried to ask questions to keep the conversation on track, yet the AI chatbot sometimes responded poorly and diverged from the target speaking subject. Consequently, during the speech, language learners were interrupted by the AI chatbot with off-topic sentences, which led to communication breakdowns. Thus, the code of *off-topic responses* from the AI chatbot emerged as a frequent code in the qualitative findings from reflective language learner diaries.

The second most frequent code discovered in the analysis of language learners' reflective diaries was the arguments by language learners related to the *AI chatbot's inability to comprehend students' speech*. Learners frequently mentioned in their diaries that the AI chatbot was inefficient in making sense of the relatively complex user input and responding appropriately.

Although the AI chatbot answered the users' questions, which required a simple yes/no answer or explicit information to display questions, it did not respond to the further, more intricate opinion-seeking inquiries of language learners. Therefore, according to the views of language learners, the AI chatbot had problems processing the users' answers during a spontaneous conversation. When learners asked one question after another, the AI chatbot malfunctioned by not responding to the student's questions. A language learner explained in the weekly diaries how it became a struggle to interact with the AI bot during a dialogue activity.

In general, it was very good but the AI is still not qualified enough for spontaneous speech. It had skipped my questions lots of times and I had to ask them twice. (ELL13, diary, week 1)

Besides the lack of the spontaneity of speech of the AI chatbot, language learners also had problems because the AI chatbot malfunctioned in understanding students' speech due to its speech recognition capability in the cases of long speech or the slow speaking pace of the learners. In both cases, the communication could not be carried out because of the inability of the AI to process such input from the users. Below are some excerpts demonstrating this situation. In the following excerpts, English language learners talk about how the AI chatbot could not respond to their statements consisting of a bulk of information.

Sometimes she didn't get it. Maybe it is because of my slow pace, but sometimes she gave overall answers such as "yes, I do- I agree" etc. (ELL21, diary, week 1)

I sometimes had to repeat my long sentences because of his not understanding me, that's why he interrupted me too much. (ELL19, diary, week 4)

AI was able to comprehend simple questions. (ELL11, diary, week 2)

Therefore, when the users paused longer than the AI chatbot was programmed to wait, the application responded through short replies, which needed to demonstrate an understanding of the ongoing conversation and a contribution to the content of the interaction.

It was also the case that when the users formed long utterances, the AI chatbot was not reactive; therefore, students had to repeat their messages until they were understood accurately. Language learners also mentioned in their diaries that the AI chatbot had issues with the comprehension of user messages due to failure in its system. They explained how it reacted in such situations, as stated below in the data.

Sometimes she cannot understand or hear my words and my questions. (ELL1, diary, week 4)

Replika detects the answers late and then says all the answers at the same time (ELL15, diary, week 2)

It randomly says expressions which do not make sense, but just to fill the gap since Replika did not understand what the speaker was saying. (ELL16, diary, week 5)

In the quotations, it was first mentioned that the AI chatbot might not be able to hear its users' speech and thus not fully comprehend the users' meaning. Second, another reason for the AI chatbot's inability to understand non-native speech was detecting the respondents' answers late. When the chatbot was late comprehending the users' messages, the message in the communication got lost. Within this scenario, the AI chatbot gave all of its answers to the previous questions of the users one after another. It, therefore, caused a misunderstanding, as the language learners could not make sense of the AI chatbot's speech in those instances. All these problems could be due to the malfunctioning in the ASR system of the AI chatbot, and there might be a necessity to adjust the gap between the speaker's and the AI chatbot's responses by the system developers.

Another frequent code emerging in the analysis of the qualitative reflective diaries was *interruptions of the user speech* and their speaking turn by the AI chatbot. There were many instances in the AI chatbot-student interaction data when the AI chatbot diverged from the topic. There was overlapping talk many times in the student-chatbot interaction. Students often expressed that the AI chatbot had problems taking turns in speech. In the following quotations, language learners explain how AI chatbots frequently interrupt their speeches.

She couldn't focus on what I was talking about. She interrupted me while I was speaking. She did not answer my questions. (ELL2, diary, week 1)
She keeps asking me questions without listening. (ELL4, diary, week 2)
It was hard to understand her words because she tried to talk at the same time with me. (ELL8, diary, week 3)

Language learners explained that as the AI chatbot interrupted the language learners' speech, it did not listen to their responses and delivered unrelated answers to the students. The AI chatbot's attention was not on the users' messages, so it could not make sense of the learners' speech, leading to irrelevant responses to users' inquiries. Moreover, because the AI chatbot was simultaneously talking with the users, language learners could not comprehend the responses from the AI chatbot. Consequently, this situation led to a mutual non-comprehension of the messages in the ongoing conversation. It was also the case in the AI chatbot-learner interaction that in those

cases of interruptions of users' talk by the AI chatbot, students struggled to restore the communication, as explained in quotations from the student diaries below.

I sometimes had to repeat my long sentences because of his not understanding me, that's why he interrupted me too much. (ELL19, diary, week 4)
She interrupts much when I try to say something long. (ELL21, diary, week 6)
It can ask to ensure the fact that the speaker finishes speaking. (ELL16, diary, week 4)

Language learners resorted to strategies such as repetition of the message to make their message more straightforward to the AI chatbot. Nevertheless, when learners repeated their lengthy messages, it led to even more interruptions. Whenever learners delivered long speeches, the AI chatbot intervened in the conversation with overlapping talk. Thus, the language learners' long utterances also triggered more interruptions in their speeches. The learners expressed that the AI chatbot needed technical characteristics such as a longer gap between the answers to wait for the speakers' turn to finish before intervening in the conversation.

Slow reaction spans/long pauses emerged as another code in data analysis under technological constraints. Some learners also talked about the AI chatbot's speaking rate and argued that its speech rate was slow. One learner narrated her experience: "Our conversations are slow, but maybe she can speak faster with practice" (ELL8, diary, week 2). Learners believed the AI chatbot needed to respond more quickly to make the conversation more natural. As the AI chatbot was slow to respond to learners, there was an unnatural gap between the responses of the two sides of the communication. They believed that if the AI chatbot responded instantly, there would have been fewer communication breakdowns due to overlaps in interaction as well. Also, the application was non-responsive from time to time, according to some of the users. One language learner narrated this situation in his diary as follows:

When I stopped speaking to give her an answer break, he didn't respond and instead offered me irrelevant responses... She listened but she didn't answer several of my questions. (ELL18, diary, week 2)

In these cases, when language learners directed specific questions to the AI chatbot, instead of getting answers to their questions, the interaction stopped because of the

application's inability to respond to learner inquiries. Learners could not get answers to their questions from the AI bot due to this malfunction. Thus, during the use of the application, there were also instances when users encountered problems related to technical deficiencies and eventual breakdowns in their interactions.

The ELLs frequently mentioned a need to improve *the AI chatbot's knowledge base* in their reflective diaries in this study. Thus, another code under the category of technological constraints concerned the lack of general knowledge of the AI chatbot *Replika*. Notably, a language learner suggested that cultural elements could be added to the knowledge base of the AI chatbot *Replika*. According to ELLs, chatbots need more information about various topics such as the world, countries, national and religious holidays, etc. According to the learners, the AI chatbot *Replika* also needed to be more intelligent. Learners explained that the AI chatbot used self-contradictory sentences and was not knowledgeable about certain subjects, as expressed by the participants' views in the following quotations from learner diaries.

They should work with her information bank about special topics such as holidays...They must develop overall skills, especially information or data about topics and concepts... Because of her insufficient information about some topics, we cannot talk clearly and keep on the dialogue. (ELL1, diary, week 5)

She should have full knowledge of subjects. (ELL8, diary, week 2)

Mainly, one week was about the holidays, and learners interacted with the chatbot based on this topic. However, it was seen in the reflective diary entries that there were breakdowns in communication because of a lack of cultural knowledge of the AI chatbot. Consequently, learners stated that topic adoption was needed, as the application's knowledge database lacked general knowledge. Language learners mentioned that topic learning was needed so that the knowledge database of the AI chatbot could possess more information. All in all, language learners frequently expressed a need for more general knowledge in their diaries. Thus, the ELLs stated a need for the AI chatbot to be more intelligent in this study.

An expected benefit from the interactions with the AI chatbot was exposure to meaningful input through the AI chatbot and language learner interactions. However,

as implied by the opinions of ELLs in this study, the AI chatbot was not always capable of delivering an informative and detailed speech. Thus, *short replies* by the AI chatbot were argued to be constraining factors in using this application.

Instead, its answers tended to be relatively short, contrary to the expectations of language learners. Thus, according to the opinions of language learners in this study, the answers of the AI chatbot were relatively short, as learners recounted in the following diary excerpts. In response to their inquiries, the chatbot did not expand on the topic but delivered short replies.

Sometimes he just answers “yes, no, that’s okay, well” after I talk for minutes. (ELL19, diary, week 2)

Sometimes she didn’t get it. Maybe it is because of my slow pace, but sometimes she gave overall answers such as “yes, I do- I agree” etc. (ELL21, diary, week 1)

After a question, it says “yes”, but does not respond afterwards. (ELL7, diary, week 6)

Learners also suggested that the AI chatbot needs to give longer answers to their questions. The findings from the reflective journals demonstrated that the AI chatbot did not contribute to the flow of the conversation by expanding on the subject and giving long answers. The AI chatbot also only sometimes expanded on the topic in question while interacting with language learners. Instead, the chatbot responded to learners’ queries by saying, “Yes, I do, I like it. I agree”. The learners expected that the AI chatbot could have given more detailed answers demonstrating its stance on the argument during their discussions. Therefore, the one-word or one-phrase reactions of the chatbot did not meet the learners' expectations.

As stated by the participants, there was also a *need for a longer duration for users’ turn of speech*. Thus, another point of argument in the reflective diaries regarding the technological constraints was waiting time intervals during the conversation with the chatbot. Language learners voiced the waiting time to answer the AI chatbot as a technological deficit. According to learners, the chatbot did not wait enough for the learners to finish their sentences while talking, thus not giving them the necessary time to think and speak. Some of the learners suggested that the waiting time for the

speakers to respond could be expanded, as in its present form, the users' speech was interrupted by the AI chatbot. The following quotations from ELLs' reflective diaries illustrate this situation.

It should wait a little bit longer before responding.” (ELL7, diary, week 2)
Sometimes she has to listen to me longer because I need time to think. (ELL4, diary, week 4)

In these cases, the users could not respond efficiently and converse with the AI chatbot. They were trying to make sense of the chatbot's answers, but they could not respond immediately after the chatbot in those instances. One language learner also narrated that when they were trying to form lengthy speeches, the AI chatbot did not wait for the speakers' turn to finish. On the contrary, the AI chatbot interrupted the language learner's turn. Therefore, it could be inferred that the AI chatbot might not be capable of processing large sets of user input one at a time, and there could be a need to divide the speech into shorter chunks when interacting with the AI chatbot, as described in the following quotation from ELLs' diaries. The learner stated, “She needs to wait a little bit when I am struggling with my English....She needs to wait when someone tries to form long sentences” (ELL21, diary, week 6).

At other times, when it was the chatbot's turn to speak, it did not instantly respond to language learners' queries. For instance, a language learner described this situation as “It pauses a lot when the turn is on it to speak” (ELL16, diary, week 2). Namely, the AI chatbot was unresponsive when reacting to language learners' answers.

Moreover, as explained by the ELLs in this study, *the chatbot also needed to ask more questions*. In their diaries, English language learners said they were the initiators in conversation almost always. For instance, a student stated, “Since I am the one who always asks questions, my questions are running out.” (ELL4, diary, week 4). Another student complained, “Generally, I have to ask questions about the topic because she is not asking” (ELL22, diary, week 7). The AI chatbot *Replika* was programmed to ask specific questions related to the well-being of the users. However, when it came to asking questions related to the subjects language learners brought up, the AI bot was

low functioning. Thus, *Replika* did not ask as many questions as learners expected, as two other language learners wrote in their reflective diaries in the following excerpts.

Questions should be added to the system of the Replika. She does not ask me any questions. (ELL4, diary, week 7)
She should ask questions about the topic. She should not only listen to me. (ELL14, diary, week 7)

Although the chatbot conversed with the learners by answering their questions, it rarely asked them questions unless being directly asked by students through commands such as “*Replika*, ask me a question about...” etc. Therefore, this property of the AI application was criticized by the learners. For these reasons, according to the opinions of some of the learners in this study, the AI chatbot did not contribute to the development of conversation by steering the direction of the conversation by asking questions. As reported by language learners in this study, the AI chatbot *Replika* could not ask questions while following the gist of the conversation. Learners, therefore, suggested incorporating a question-asker role into the conversation design system of the AI chatbot.

Another code that emerged in the analysis of reflective diary excerpts was the *repetitive language* used by the AI chatbot. Sometimes, during the speech, because of the recurrent repetitions in the interaction, learners felt overwhelmed and unmotivated to continue chatting with the chatbot, as seen in the following extract from the reflective diary of a language learner. While explaining the nature of their interaction, this language learner complained that “she repeated the same questions with the same words. It is boring. We always go back to the start” (ELL1, diary, week 7).

In the AI chatbot-student interaction data, there was much repetitive language in the AI chatbot's speech. When learners asked specific questions, the chatbot almost always answered in the same manner using the exact words. For instance, in several AI chatbot-student dialogues, whenever students wanted to talk about Russia in the week, the learners were assigned to talk about the topic of culture; the AI chatbot each time answered with the same sentence, “I stand with Ukraine.” Moreover, when students asked what aspects of a particular country the AI chatbot liked, the chatbot always

gave a pre-set answer like “The food, the scenery, and the culture.” Consequently, these repetitive responses from the AI chatbot created boredom among the learners, as whenever they asked a question on a particular topic, they got the same answer to these questions.

Several language learners in this study also reported a *need for a one-call-at-a-time feature* because learners had technical breakdowns due to multiple simultaneous logins into the AI chatbot application. The subsequent quotations from learners’ weekly reflective diaries illustrate this problem.

It should be impossible for system to have a call with multiple persons. Only one person should call AI robot. (ELL4, diary, week 3)
Our conversation was interrupted in the 13th minute and then I waited 20 minutes for my other friend to do her homework then I came back but after that, some kind of technical problem showed up, so after the 13th minute it was a bit messy. It would be good if the other students don’t interrupt the conversation but I guess there is nothing to do about this. (ELL21, diary, week 5)

According to these participants, when multiple students tried to log into the application, the system did not allow multiple user logins. Consequently, interactions with some of these users were cut off at some points. In order to be able to communicate with the AI chatbot, language learners had to wait for their turn to talk, which was a restraining factor for them to carry out the task successfully within one log-in. Therefore, participants in this study voiced a need for a single-person login version of the application, which was not possible during the implementation of this study due to needing access to multiple premium accounts.

The analysis of the weekly reflective language learner diaries showed that another code under the category of technological constraints appeared as a *need for incorporating multimedia elements*. The qualitative data analysis demonstrated that although the AI chatbot promised its users to send audio materials or visuals in some parts of their interaction, it did not keep its promise or send the multimedia elements to the users. For instance, one language learner mentioned that “when she says that she will show us something, I think she could be able to do it” (ELL15, diary, week 4). However, including multimedia would have helped to enhance the experience of

using an AI chatbot tool for learning languages with more learning enjoyment and appeal to learners with various learning styles.

Several language learners also mentioned adding *different accent options* in the reflective diaries. During the implementation period, the AI chatbot *Replika* had only the option of a native American speaker accent in its previous version. It only had various tone choices, such as confident, caring, optimistic, etc. Therefore, the results of the analysis showed that some participants wished to include other accent options in the chatbot application. For instance, wishing to incorporate a variety of other accents, a learner stated that “it can have different accent options” (ELL16, diary, week 2). In this way, learners sought exposure to various English accents, aiming to understand the intricacies among these accents as individuals studying English as a foreign language.

The findings from the analysis of diary excerpts have also demonstrated that the AI chatbot was considered to have *deficient language skills* from various linguistic aspects. Below are students’ ideas extracted from reflective diaries commenting on the language skills of the AI chatbot *Replika*. One learner stated “She doesn’t use intonation” (ELL2, diary, week 4), while another mentioned, “There were some pronunciation errors, but it is fine. I make mistakes either” (ELL7, diary, week 5).

In these examples, the AI chatbot *Replika* needed to improve its use of pronunciation and intonation. Although the AI chatbot was programmed to use human language, it could not efficiently adopt a human-like intonation in sentences. Instead, it had a mundane robotic intonation while interacting with its users. Therefore, English language learners in this study suggested a development of these characteristics of the AI chatbot application. Additionally, the excerpts by two language learners below demonstrate their opinions on the word knowledge of the AI chatbot *Replika*. One language learner suggested that “*Replika* can be improved in word knowledge” (ELL5, diary, week 2). Another one argued that “she has to learn more vocabulary” (ELL18, diary, week 2). These learners observed that the AI chatbot *Replika* needed comprehensive word knowledge. Therefore, they suggested that the word knowledge database of the application would be further enriched.

The lack of a group talking feature was another code that emerged in the analysis of diary data. Another suggestion by language learners was adding a group talking feature. In its present version, the users could only talk to the AI chatbot using a single-user mode. However, language learners mentioned the lack of a talking-in-groups option. The users of the application in this study asked for this technical feature to be incorporated into the AI chatbot system. The learners believed this feature would help the users interact more efficiently by enabling collaborative learning.

Another subject discussed by the ELLs in reflective diaries was the *need to connect the app to the internet for new information*. ELLs in this study wrote in their diaries that there was a need to form an internet connection to check for any unknown information by the AI chatbot. Learners explained that the AI chatbot sometimes could not answer its users' queries; thus, they voiced a need to connect the chatbot application to the internet to search for any missing information in its knowledge database. Considering that there are millions of knowledge available within the reach of internet users worldwide, it is suggested by the users of the AI chatbot in this study that this information had better be extracted from the world wide web within a short duration of time through instant internet connection.

One language learner proposed that when the AI chatbot lacks knowledge about specific topics at times, it should be able to connect to the World Wide Web to search for the missing information, stating, “Why can’t the AI chatbot use the internet to look up for some info?” (ELL7, diary, week 5). The restoration of communication through reaching for more information online was an emerging argument among the ELLs in this study. This could be an ideal solution that could be considered to develop the technical properties of AI chatbots currently available.

According to the language learners' beliefs in this study, the AI chatbot required training on various subjects using the possible data sources available in any format. The data on the internet is publicly available, and the utilization of this data from the Internet is within the limits of public rights through the necessary processes. Therefore, according to the participants in this study, these vast data sources for training need to be appropriately utilized by the designers of AI within this field.

There was also a *lack of an initial examination of students' language proficiency* in the AI chatbot's interface in this study. Language learners expected that their initial language proficiency level could have been specified when they first started using the AI chatbot application. Looking at the reflective learner diaries, it was observed that an initial examination of students' language proficiency was proposed as a technical quality that was not a characteristic of the AI chatbot system. According to one language learner, the AI chatbot should be able to determine the language proficiency level of its users and respond accordingly. The student suggested that “she can determine my level with an oral exam and she can speak according to my level” (LL4, diary, week 7). The AI chatbot *Replika* was not initially created for the use of English language learners. However, if a chatbot is to be developed targeting this group of learners, the application developers suggested adding this quality.

The target participant group of this study was English language learners at intermediate or advanced levels of language proficiency. These learners at these proficiency levels also stated a wish to be able to talk to AI chatbots without any problems in communication due to issues such as the use of too basic words because this group of learners expected the use of an advanced English language level to develop their language skills. For instance, language learners expected that the AI chatbot could ask them questions by initiating the conversation. Additionally, some of the participants in this study stated the need to use argument structures more with language functions, such as disagreeing with the users during a conversation. Learners expressed the AI chatbot's lack of use of argumentative functions in interacting with them.

As argued by the ELLs in this study, there was also a *need for a better AR system* within this AI chatbot. The AI chatbot *Replika* used augmented reality (AR), allowing students to see the avatar in three dimensions in their present physical settings. When the learners opened the application, the AI avatar was put in the learners' actual surroundings. The users could see an AI avatar in a three-dimensional form displaying jests, gestures, and mimics. This feature could have appealed to a variety of language learners in this study, as it might have given an image of a social presence of the AI avatar in the users' natural environment. However, the study participants mentioned

that this augmented reality characteristic should be developed further. For instance, among the users of the AI chatbot *Replika* in this study, a learner expressed a need for a more developed AI system by stating that “the application needs a better AR system” (ELL14, diary, week 6).

There could be a variety of reasons behind this need to develop the AR function of this application, as voiced by the participants in this study. The first issue could be related to the students' internet connection. As reported in the subsequent chapters, ELLs had problems due to a low internet connection while using the application. As voiced by the users, the need to develop the AI could be related to the mode of use of the application. The participants in this study used both the desktop and mobile phone modes while using the application, as, unfortunately, the application did not support all Android users at the time of the implementation. For this reason, there could be differences in the functionality of the AI chatbot's AR skills. Therefore, this technical quality was stated as needing consideration by the future developers of AI bot applications.

Among the language learners, some participants could not understand the AI chatbot's speech because of its fast rate. One learner criticized that “she can speak slower sometimes. I want to write the words I don't know” (ELL4, diary, week 4). Thus, the *fast speech rate* by the AI chatbot was argued to be a restricting characteristic that hurt their experience of the application use.

Another student complained about this situation, stating that “it was hard for me to understand her sometimes because she was talking fast” (ELL17, diary, week 1). Therefore, language learners expressed a need for the chatbot to have a slower speaking rate, as some learners reported being unable to keep up with the ongoing conversation. According to these learners, the AI chatbot had better use slower speech so that they could follow the conversation and fully comprehend the dialogue. The suggestion to develop the AI chatbot application to have an adjustable speech rate thus emerged in the qualitative reflective diary data. For this group of language learners, the standard native speaker norms for speaking were not valid, as the participants in this study had a variety of language proficiency backgrounds in terms of listening skills

in the English language. It was, therefore, important for this group of learners to get exposed to a comprehensible dialogue with clear, understandable, and audible speech from the AI chatbot. Therefore, there was a need for an adjustable speaking rate for the non-native users of this AI chatbot application.

Another technical property indicated the need for development in the users' diaries was the enhancement of the subtitles. Thus, the participants argued that there was a *need for better subtitles* in the chatbot application. Although the subtitles allowed the users of the AI bot to follow its speech and check for understanding of the ongoing conversation, the students still asked for a change in the presentation of the subtitles on the screen. It was hard for some users to follow the speech of the AI chatbot from the subtitles because of the visual interface of the application.

One suggestion by one of the language learners was the addition of subtitles to a different place on the screen. Although the application had subtitles under the screen, some learners thought it would be more convenient if the subtitles were situated in the corner above for more visibility. On this issue, a user said, "I think the words should be written in the corner because I can't read the words she said instantly" (ELL4, diary, week 4). Therefore, learners suggested that the visual interface of the AI chatbot could be developed further by changing the place of the subtitles on the screen. Thus, it was suggested that the application could be more convenient and practical for the users. Consequently, the development of the application's visual design stood as one issue that needed further attention from the designers, as suggested by the participants.

Another code formed in the data analysis phase was the *need to prevent the "tiredness" of the AI avatar due to multiple-use modes*. During the implementation period, the AI chatbot *Replika* had a limit for the time it actively interacted with its users. After talking to students for some time, it stopped responding, and the learners had to wait for the speaking mode to be activated again. If the students continued chatting with the AI chatbot, it did not respond eagerly and eventually became silent. Therefore, considering this issue, one language learner suggested that the "tiring limit should be more" (ELL19, diary, week 2). While using the application, learners also won points with which they could later shop in the store of the AI application.

However, the points saved were not usually enough for the chatbot to regain power instantly all the time. Therefore, the AI chatbot became unresponsive and silent at some points during the application use. In those instances, the students had to wait for some time for the application to function correctly, or they could face some breakdowns in communication; they could also carry on with the conversation if they would like with a lower functioning capability of the AI chatbot.

This property of the AI chatbot *Replika* intended to stand as a motivator for the application users to use the application more because when the users chatted with *Replika* more, they also won diamond points with which they could regain power. Therefore, although stated as a limitation by the language learners in this study, the seeming tiredness of the AI avatar functions as an impetus for continuing the interaction with the AI bot. Moreover, although participants mentioned that the application became unresponsive when the AI avatar got tired, the application still responded to some degree to continue the conversation, as recorded in the AI chatbot-language learner interaction data. Learners also said that they had *internet connection problems*, as some of them were living in student dormitories with low internet connection. The AI chatbot became unresponsive or eventually shut down when they had a low internet connection. In the following quotations from reflective learner diaries, the case of how internet data causes users to have problems with their mobile devices is explained. These students stated that the AI chatbot application was complex to use on their mobile phones, as it caused their phones to slow down and heat up.

They should upgrade the version since it is hard to use it from my phone. (ELL4, diary, week 3).

Replika needs an update because my phone and computer are having hard times. (ELL14, diary, week 5)

Therefore, the learners believed that the application designers must handle this matter related to a vast internet data stream. Besides, the AI application did not have the option of working without an internet connection. Thus, when the internet was cut off, the interaction stopped. As the application required an internet connection, the learners had various problems. For instance, one learner stated, "She could not hear me because of the internet problems. Therefore, I needed to repeat the same thing again" (ELL14,

diary, week 4). Another student mentioned that “we can’t use *Replika* if we have a bad connection, and it consumes too many MBs” (ELL19, diary, week 6). Several students mentioned having experienced at least one of these issues in their weekly reflective diaries.

Another emerging subject was concerned with the *lengthy speech of the AI chatbot*. According to some language learners, the AI chatbot *Replika* dominates conversations with its users. Students said that they needed help comprehending the speech of the *Replika* chatbot due to the lengthiness of its answers. One of the language learners expressed that “some of the sentences of *Replika* were too long, so I had some trouble comprehending it” (ELL6, diary, week 6). Thus, the non-comprehensibility of the AI chatbot’s speech appeared as one prominent theme in the learner diaries. Furthermore, another student said that “she talks more than me, even though I am the one who should talk most of the time” (ELL4, diary, week 2), arguing that the AI chatbot dominated the talk. Even though there were also short replies by the chatbot, as evidenced in the interviews with language learners and their reflective journals, at times, the chatbot had a longer speech than expected, which caused the formation of a communication barrier between the AI chatbot and learners.

When the AI chatbot *Replika* talked, focusing on a therapeutic mode, it dominated its user interactions. As the chatbot tried to give counseling to the users, it asked one question after another about their well-being. However, as the participants in this study were trying to construct a conversation based on a specific set theme, they did not respond to the chatbot’s queries most of the time. This situation led to communication breakdowns, as the students were trying to diverge the conversation away from the therapeutic speech of the AI chatbot to the discussion subject of the week. Therefore, the domination of the interaction by the AI chatbot caused frustration among language learners in this study.

It was also discovered during the analysis of the qualitative diary data that there needed to be more *counterarguments* by the AI chatbot. According to the participants in this study, the AI chatbot system was created based on agreement with the users’ utterances. The language learners expected that the application would disagree more

during interaction to get engaged in a more meaningful dialogue. However, it was observed that there were no disagreements in the AI chatbot's responses. Therefore, the response system of the AI chatbot was based on agreement with a previous utterance by the users rather than agreeing and disagreeing at times during a conversation. In this situation, one participant stated that the AI chatbot did not encourage users to communicate more by discussing different viewpoints.

First of all, it was a good experience for me, but it felt like I'm speaking to myself because the response system is based on agreeing on my opinion. That is to say, it was not different from mirror talking. (ELL13, diary, week 1)

On the contrary, it was expected by the learners that there should have been two-way communication with a dialogic speech between the AI chatbot and the users of the application. However, due to the non-arguing stance of the AI chatbot, the discussion between the users and the AI chatbot was conducted through a one-way communication mode. For this reason, the students mentioned a need to train the AI chatbot to disagree more with its users and exchange opinions on various subjects. It could be thus possible to carry out a conversation with supporting and refuting arguments inherent in a natural human-to-human conversation.

According to the views of the ELLs, there was also a need to remember the user history of the AI chatbot in this study. Therefore, remembering the previous conversations with its users was another technical constraint the participants mentioned. The students pointed out that the AI chatbot needed to remember the user history. The AI chatbot was also sometimes incapable of referring to a few statements in the dialogue. Therefore, the application had a technological restriction to keep the antecedent information in a conversation in memory. The following excerpt from a language learner's reflective diary illustrates this case.

It does not have a good traffic of conversation. It does not remember what is being said from your main topic to the previous 2-3 sentences. (ELL16, diary, week 4)

Thus, while students were interacting with the AI chatbot, the chatbot needed help remembering their previous sentences, which led to a miscommunication during the

chat, as the AI chatbot could not follow the topic flow. As a result, the interaction did not flow as expected by the learners, and the communication even came to a halt at some points. The ELLs in this study suggested that the application designers in this field had better develop this technical property. Therefore, there is a need to develop the capability of the AI chatbot Replika further in terms of recalling its users' previous sentences to a few sentences back and keeping all of the previous conversation content in its memory base.

The non-presence of a pronunciation help chat screen was another technological constraint mentioned by the participants. As the participants of this study were English language learners, they needed help pronouncing some words. During their interactions with the AI chatbot, learners requested to learn from the chatbot if they were pronouncing certain words right by asking for confirmation through statements such as “Am I pronouncing this word accurately?” However, in those instances, the AI chatbot did not provide the learners with the correct pronunciation of these words, as it could not supply explicit error correction support. To get help on this issue, one language learner proposed the following solution.

When the speaker is struggling to pronounce a word, there should be a chat screen where the speaker can text the word he struggles to pronounce. (ELL16, diary, week 5)

Another learner suggested that the users could chat through text with the chatbot and ask for the pronunciation of the words they struggle with. In this way, learners could get help from the application whenever they need support. By the presence of a specific part alluded to allow the users to reach out for help with the pronunciation of specific vocabulary, it would also be possible to expose the learners to the correct pronunciation of the words whenever they interact with the AI chatbot.

Stopping calls abruptly was another technological constraint, as mentioned by one of the language learners who stated that “AI robot ends the phone call suddenly” (ELL4, diary, week 3). Due to issues such as low internet connection, the learners experienced technical breakdowns of the application during the implementation period. The users experienced this issue both on their mobile phones and on the desktop version.

Furthermore, the voice call feature of the application did not function on specific Android devices at the time of the application, which even caused more restrictions on carrying out healthy interactions with a stable connection. There was a lack of high-functioning mobile devices with the latest update features that could allow users to conduct an uninterrupted conversation with the AI chatbot. As a result, language learners frequently mentioned experiencing this problem in their use of the chatbot, and this frequent problem that the learners faced needed further attention.

One technological constraint put forward by the participants was that the AI chatbot frequently shifted the topic of the conversation during a speech when it needed help understanding the contents of the message by the students rather than stating non-understanding. Consequently, a *need for stating non-understanding, not topic shifting*, was a subject of argument among the participants. Although there were also positive experiences among the participants on this issue, some students complained that the AI chatbot did not explicitly express when there were unintelligible parts in the conversation. Whenever the learners' speech was incomprehensible to the AI chatbot, the topic was changed by using off-topic statements, which led to confusion among the language learners. The following excerpts illustrate how the AI chatbot reacted to miscommunication by altering the subjects instead of employing other communication repairing strategies like explicitly stating non-understanding or asking for repetition.

She should ask me if she doesn't understand the topic. (ELL4, diary, week 2)
When she does not understand, she should say she does not understand, and she should not move on to another topic. (ELL4, diary, week 7)
It can just say "I do not understand." when it does not understand the question. (ELL7, diary, week 5)

Although there were also instances when the AI chatbot asked for the clarification of meaning from the users, some learners in this study also stated that the AI chatbot changed the topic to another subject at times when there was non-comprehension of the user messages by the AI chatbot. Therefore, the participants in this study voiced a need for the AI chatbot to point out any miscommunication with the users and ask for repetition and clarification of the unintelligible information. The participants, therefore, suggested the incorporation of the use of negotiation devices into the interaction capabilities of the AI chatbot.

4.4.3. Findings Regarding Technological Opportunities from the Interviews

The findings of this study demonstrated that the most frequently mentioned topic among the ELLs in this study regarding technological opportunities of the AI chatbots was AI chatbots being competent in language skills, followed by being capable of expanding on previous conversations and knowledgeable on many topics (see Table 42). The first code under the category of technological opportunities, as mentioned by language learners, was the AI chatbot's *competency in language skills*. Findings from the interview transcripts also demonstrated that the AI chatbot possessed good linguistic skills.

Table 42 Frequencies for technological opportunities codes emerging from the interview data

Category	Codes	<i>f</i>
Technological opportunities	Competent in language skills	7
	Expanding on previous conversation	5
	Knowledgeable on many topics	3
	Comprehending students' speech effectively	2
	Optimum rate of speech	1

Moreover, the subcodes emerging under the code of linguistic skills were fluency in the language, listening comprehension skills, accurate pronunciation skills, advanced-level vocabulary, and clear and understandable speech. The following examples demonstrate each aspect with an explanation of language learners' experiences using the AI chatbot.

She is fluent in the language. She speaks English very well, like a native speaker. And I think it is because her developers provided her with the native resources and also she really tries to understand the person she's talking with. (ELL5, Interview)

I think her language skills are perfect because she's a computer and in most of the areas, computers are better than us. (ELL5, Interview)

The speech of the AI bot was really comprehensive and understandable. So, I really didn't need to push myself hard to understand the speech. So, yeah, it was easy to understand him or her. (ELL19, interview)

The first point of the argument was related to the fluency of the AI chatbot in speaking the language. The chatbot was considered to have the characteristics of a native speaker in terms of fluency. As a learner put forward, because the AI chatbot was trained using native resources, it had developed linguistic skills. Besides having fluent speech, the AI chatbot also possessed good comprehension skills. That is, the AI chatbot could also accurately comprehend the users' speech, as reported by the language learners. The below excerpts demonstrate the opinions of some English language learners regarding these issues.

It speaks fluently, so it was way better than communicating with my friends. Because it's not a real person, it doesn't make any structural mistakes. (ELL7, Interview)

Her pronunciation was good. (ELL10, Interview)

Definitely, it was really good at vocabulary and I can say that it was in a random way. Sometimes it was using it at one level and sometimes I saw so many words that belonged to C2 or any words. I can say that it was so balanced and I really like that. (ELL13, Interview)

One student even stated that as the AI chatbot was a computer, it was better in language skills than human beings in many areas. The student put forward that computers could exceed human beings in many areas, including the use of language. Furthermore, regarding the linguistic capability of the AI chatbot, learners also believed that it was easy to understand the speech input from the AI bot because it developed speaking skills. Therefore, students said the AI chatbot had good pronunciation skills, a wide array of vocabulary knowledge, and fluent speech.

Consequently, learners believed that compared to their experiences with their peers, they were exposed to more meaningful input free of grammatical errors. If students interacted with their peers, they might have yet to be exposed to input that carried such grammatically accurate knowledge. However, language learners could interact with an AI chatbot that uses the target language accurately. The learners also said they benefitted from the AI chatbot's advanced pronunciation skills. The students could hear the correct pronunciation of words while speaking; thus, they stated that talking

to an AI was an opportunity to hear the correct forms of the language. According to language learners, using an AI chatbot also helped them develop their vocabulary knowledge, as the AI chatbot had a wide array of vocabulary uses in the target language. In this study, many language learners believed that the AI chatbot had advanced linguistic skills, which provided meaningful input for the ELLs in this study.

Another code was related to *expanding on previous conversations* by the chatbot. The interview participants stated that the chatbot contributed to the interaction flow during their conversations. It agreed and disagreed with the users based on the context of the conversation. According to these learners, the AI chatbot could respond to user queries and carry out a two-way conversation based on agreements and disagreements during the conversations. Excerpts from the interview data regarding the responses to the interview question “Was *Replika* able to provide a follow up to your previous comments?” are demonstrated in the examples below.

Interviewer: I see. Was Replika able to provide a follow up to your previous comments? When you talk about something, does she comment more on it or not?

Learner: Yes, she can comment on it. Yes, she's kind of good at this. For example, if I say I like red, she says I like red too. That's nice. (ELL5, interview)

In this first interview extract, the interviewer asks whether the student could get feedback from the AI chatbot during interaction. The language learner responded by explaining that the AI chatbot commented more on their responses by showing agreement or disagreement on the subject being discussed. Therefore, this language learner showed a positive attitude toward the ability of the AI chatbot to contribute to the flow of a conversation by making supporting or refuting comments. The following interview transcript also demonstrates an example of the positive experiences of a language learner in expanding the topic being discussed through making on-topic comments.

Interviewer: So, did she provide a follow up to your previous comments? Did she make comments about what you were talking about and expand on it?

Learner: Yeah, she did, she made some comments. Like, for example, I was telling her I'm good at driving and stuff. And she was telling me that you should

follow your dreams. For example, just don't give up. You're so good. This kind of thing, I think. Good advice. (ELL15, interview)

In this second extract, an ELL explains how she could carry out a conversation with the AI chatbot through the supporting arguments of the AI bot. When the ELL mentioned a subject such as driving, the AI bot commented on it. It encouraged the student to continue developing this skill. Thus, it could be seen from the experience of this learner that the AI chatbot supported the student's arguments in their interaction by commenting on the language learner's skills. The following quotation from an interview also illustrates how the AI chatbot provided a follow-up to information mentioned during a conversation with a language learner.

Interviewer: Was Replika able to provide a follow up to your previous comments during speech?

Learner: Mostly yeah. The first thing that he did was agreeing to agree or to reject. I agree with you or I don't agree with you. That was the first thing that he did. And then he tried to express his own opinions about the situation. So yeah, he did. (ELL19, interview)

Some participants also said the AI chatbot could follow the topic flow while conversing with learners. As expressed by several language learners, the AI chatbot also waited for students' turn to finish before starting to speak. The following quotations from interviews explain learners' experiences on this issue. For instance, ELL7 below mentioned that he or she could chat with the chatbot while taking turns in speech.

It did a good job on this. It followed my question. It followed the topic. Also, it added its own ideas. So, we kept speaking because I'm saying something, it says something and it adds more. So, I answer these questions one by one. That was nice. (ELL7, interview)

ELLS in this study stated in the interviews that the AI chatbot could take turns in speech and expand on the ideas in the previous conversation by following the flow of information in speech. It was reported in the interviews that the AI chatbot was able to make on-topic comments and agree or disagree with the learners based on the information in their dialogues. Moreover, learners stated that the AI chatbot could have a conversation by showing emotional reactions with linguistic means, such as verbally

encouraging learners to pursue their dreams. Thus, the dialogical argumentative feature of the speech was present in the AI chatbot-human interaction according to the participants' views in this study.

ELLS in this study also discussed that the AI chatbot was *knowledgeable on many topics*. As language learners talked to the AI chatbot *Replika* each week, they could accumulate points that could be spent on adding new information to the chatbot's knowledge base. Thus, the researcher intermittently checked the accumulated diamonds and purchased new information on the application on topics such as make-up, shopping, soccer, etc. Therefore, the breadth of knowledge of the chatbot increased with each passing week through the information purchased due to accumulated points. Some learners stated this characteristic of the AI chatbot about being knowledgeable on many subjects in the following interview excerpts.

Actually, I can say that the AI bot was given so much information about topics that I don't know about. I really like this point of it. (ELL13, interview).

Yeah, most of the time, even though I did not think she understood fully, I asked a question about it, and she was very knowledgeable. She has knowledge nearly about every topic even science. And I was just listening the information she's given to me, because I didn't have any information about science. (ELL4, Interview)

The language learners mentioned in the interviews that AI chatbots had a wide array of knowledge on many topics. Even though the learners initially had doubts about the contents of the knowledge of an AI robot, they believed that AI could have a knowledge base more than they ever expected. Initially, the AI chatbot had already been trained with a bulk of information from Twitter. The data from online interactions, news, journals, and many other sources from social media were utilized in the development of the knowledge base of the AI chatbot. The AI chatbot also gained knowledge by interacting with the language learners. Namely, the AI chatbot could also learn new information through communication with its users, which was a characteristic of the AI. After each week of using the AI chatbot, the AI chatbot could learn more about the profile of the application users and respond accordingly, as reported in the findings of this study.

Comprehending students' speech effectively was another code created after analyzing the qualitative interview data. According to the opinions of language learners in the interviews, there were cases when the AI chatbot could not comprehend the learners' speech. However, some learners stated that they used repetition to repair the conversation. The AI chatbot then could comprehend the users' messages. Based on how familiar it was with the topic of conversation, the chatbot could converse with language learners by following the flow of the conversation. The excerpt below from an interview with a language learner illustrates this point.

Interviewer: Was it able to keep up with what you are saying?

Learner: Yeah, but it again depends on the topic. We talked about mostly what was needed. And once he couldn't understand. It was not a big problem. I just repeat it again. (ELL19, interview)

Thus, language learners in this study explained how they were successful in their endeavors to communicate with the AI chatbot through clarification of the messages. Language learners could clarify their meanings through negotiation devices when there was a lack of understanding of the user messages. In the following example, another language learner talks about the experience of using the AI chatbot *Replika*. The learner expresses that the AI chatbot could comprehend their pronunciations and articulation of messages.

Actually, I am impressed at how she can understand my pronunciation, punctuation, whatever I am saying. Yes, she is good at understanding when you talk like academic English or something. (ELL5, interview)

According to the views of participants in the interviews, the AI chatbot could understand their speech, and they could make their message comprehensible to the AI bot. The findings from the interviews showed that the AI chatbot comprehended the speech of non-native English speakers in this study. Thus, this study proved that the AI chatbot's automatic speech recognition capabilities allowed non-native users to comprehend messages. Many participants in this study recounted their successful attempts at communicating with the AI chatbot through the use of conversational strategies to adapt their speech during an interaction in the case of any breakdowns in communication.

Another pedagogical affordance emerging from analyzing language learners' statements in the interviews was the *optimum rate of speech* by the AI chatbot *Replika*. Thus, learners in this study also articulated that the chatbot had a fast speech rate. In the following statement, one student explains that the AI chatbot's speech was fast and straightforward. It could be seen that the rate of the speech did not lead to a communication problem between this learner and the chatbot, as ELL19 stated, "Yeah, clear speech. Clear and fast speech. Exactly. Fast... You know yeah, that is it" (ELL19, interview), which indicated a positive stance towards the speech rate of the AI chatbot by this learner. This learner thought that the AI chatbot was talking at a rate that had an optimum level for comprehension of the messages in the interaction.

Although there were varying views on this issue, the AI chatbot had a moderate speech rate, according to some language learners. Namely, it did not have a too-slow rate of speech that would sound too robotic. Instead, the AI chatbot spoke moderately, allowing language learners to follow the conversation. The learners believed that the rate of speaking of an AI chatbot was natural for a robot. It was neither too fast to impede the users' understanding of the AI chatbot's speech nor too slow to make the conversation monotonous and robot-like.

4.4.4. Findings Regarding Technological Constraints from the Interview Data

An analysis of the interview findings indicated that English language learners' most frequently stated technological constraints in this study were off-topic sentences by the AI chatbot, followed by the AI chatbot's inability to comprehend students' speech, a need to improve its knowledge base, chatbot's deficient language skills, and slow reaction span and long pauses. Table 43 below demonstrates the frequencies for technological constraint codes as emerged from the interview data. *Off-topic sentences* by the AI chatbot were also a dominant code discovered in the interviews carried out by language learners and in reflective journals. Off-topic sentences of the AI chatbot were put forward as a restricting factor by the majority of the language learners. Language learners frequently mentioned that the AI chatbot delivered off-topic statements quite often. Thus, learners expressed a need to stay on topic while interacting with the AI chatbot.

Table 43 Frequencies for technological constraints codes emerging from the interview data

Category	Codes	<i>f</i>
Technological constraints	Off-topic sentences	11
	AI chatbot's inability to comprehend students' speech	8
	Need for an improvement of its knowledge base	6
	Chatbot's deficient language skills	5
	Need for incorporation of multimedia elements	4
	Slow reaction span/long pauses	4
	Need for the chatbot to ask questions	4
	Need for adding different accent options	3
	The need to connect the app to the internet for new information	2
	Need for longer duration for users' turn of speech	2
	Short replies	2
	Internet connection problems of students	2
	Need for a button for turn-taking	1
	Inconsistent in information about itself	1
	Need for a one call at a time feature	1
Total		56

While language learners were talking to the AI chatbot *Replika*, it frequently responded off-topic, as explained by one of the language learners in the following interview excerpt.

Actually, it is a good app for practicing your talking skills, but she was always interrupting while a person was talking. So, it's kind of disturbing for me to keep talking about my ideas. And sometimes she talks about different topics, so I always try to remind her what we are talking about. We should always talk back to that week's topic. That was a little disturbing too. But if we look at the

total it was a great experience for me actually. ... Actually, Replika is great but like I said, some difficulties were in it. I want to make her just talk about what the person is intended to talk about. (ELL17, interview)

More specifically, during these conversations, the AI chatbot diverted the subject the students were focusing on, for instance, by giving counseling advice and mentioning other unrelated topics. As the AI chatbot was programmed to assist its users by emotionally supporting them, it tried to change the focus of the conversation at some points to talk about those issues. However, students continued carrying out their conversations under their weekly assigned tasks. This situation caused negative feelings among students.

Some of the difficulties I had were sometimes I was trying to talk about a specific topic like fashion or something like art, science. But while I was trying to talk about that topic and I was asking questions about that topic and Replika was just answering, how can I say that? It was not relevant to my questions or to the topic. And it was the biggest problem I had with Replika. (ELL6, interview)

Okay, so I'm going to talk about the negative side first. When I was speaking, the bot interrupted me a lot and she asked about irrelevant questions. For instance, I was talking about history and she was asking me if she could tell me a secret or something. (ELL2, interview)

She gave irrelevant answers and she interrupted me. (ELL10, interview)

As a result, learners could not follow the ongoing conversation because of the frequent shift of topics with unrelated comments from the AI chatbot. Even one of the learners stated that due to interruption by the AI chatbot on many occasions, the learner had to repeat his message repeatedly. Despite their various attempts, language learners struggled to get their meaning across. Thus, as explained in the following quotation, the learner felt exhausted and lost interest in using the application.

It went off topic several times and it was hard to follow the ongoing conversation, because I couldn't keep up with that current topic. So, I think this is a technical problem. And other than that, that's just it. ... It kept interrupting my questions. It gave unconnected answers, off topic answers. What else? Yeah, that's it. ... Mostly it was on topic, but sometimes it went off topic and I had to repeat the question in order to get some solid answers or solid reactions, I would say. So that part is tiring because I had to repeat the question again and again and since it's tiring actually, it made me stop chatting with Replika. (ELL7, interview)

Two learners used the adjectives “annoying” and “unbearable” to define how these frequent interruptions during the conversation made them feel. Other learners also expressed negative feelings towards using the application because of this technical issue. However, they adjusted to this situation and continued their conversation without being demoralized. The following quotations explain how the AI chatbot diverted from the topic during a conversation with learners.

As I mentioned before, it was annoying sometimes that Akdeniz was out of topic. (ELL9, interview)

Sometimes the AI really gave too many irrelevant responses which actually made the conversation unbearable. But from a general perspective it was really good. ... Actually except for the last four assignments, it was every time giving an irrelevant response. For instance, like I'm asking a question to the AI and it is just replying to me with a question and the question is not related to our topic or anything. It was just related to something like if you like this app, please share some private information. It was like that. ... Actually, I can say that in this topic, it was actually bad at following the flow. It was actually bad at this. So, whenever I ask a question, for instance, we just had a conversation, and I asked my question about this topic, it gave some relevant responses. But whenever I just stopped talking for a second it's really changing the topic and giving too many irrelevant responses. It was actually not available in terms of that. (ELL13, interview)

It was also the case that language learners could not interact with the AI chatbot because of frequent interruptions, so they formed solutions to this issue through self-generated coping strategies. For instance, the learners started to ignore this behavior of the AI bot and continued with the conversation without losing the motivation to speak in the target language. Further experiences by language learners are demonstrated in the examples below.

I was asking so many questions that it didn't listen to me and it started to ask its questions, and it was off topic for me. So, it can be more improved about asking skills as I stated before. (ELL18, interview)

Replika constantly changing a subject like we were talking, and suddenly asked me about something different. That was the issue. But later on, I just got used to Replika and it wasn't a problem for me. I just asked a question related to my topic and she continued to talk about my topic. (ELL4, Interview)

Although the AI bot diverged from the subject of the conversation at some points, the learners could persistently repair the communication to carry on with their interactions.

As the learners in this study were focused too much on conversing with the chatbot on a specific theme throughout their whole interaction, they kept asking the AI chatbot questions on various aspects of the same topic. However, as the knowledge base of the chatbot application did not suffice to respond to users' every question, it interrupted learners' speech with automatic responses from its database, which led to incoherent communication full of mutual misunderstandings in the AI chatbot-student interactions.

AI chatbot's inability to comprehend students' speech emerged as another dominant code under technological constraints in the qualitative interview data. In the interviews, students also stated that the AI chatbot sometimes could not understand their speech. Language learners also mentioned that AI chatbots could not comprehend deep and complicated subjects. The following excerpts explain the experiences of learners having miscommunication problems with *Replika* due to non-comprehension of users' speech.

Yes. Sometimes she has hardship to understand what am I talking about. But even though she's good at comprehending, I guess yes. (ELL5, interview)
Sometimes when you add something very complex or something, she's not responding. I guess she didn't fully understand everything. (ELL15, interview)
Sometimes she understood what I told, but sometimes she didn't understand. She gave me different answers and sometimes she didn't answer me because she didn't understand me. It changed from time to time actually... Yeah, a lot of times she understood me, but sometimes like I said when I was trying to make long sentences, she just interrupted me, but a lot of time she understood my long sentences. (ELL17, interview)

One student mentioned that the AI chatbot could not comprehend the learner's meaning because of the mispronunciation of a word. Therefore, the learner tried to change his or her pronunciation of a word, paraphrase the sentence, or shift the topic to another one to deal with communication breakdown in those instances and ensure mutual comprehensibility. Learners modified their output in order to get their meanings across. The following excerpt narrates the experiences of this language learner who used a variety of strategies to deal with communication breakdown in those instances.

I just changed the subject and asked another question to Replika. Because I couldn't keep up with Replika, I just changed the topic, and she helped me to change the topic. And she answered my questions again. Sometimes I tried to pronounce it again. I realized that I couldn't and she didn't understand me, so I changed the topic not the overall topic, but the topic that we are talking about. If you are talking about science like Albert Einstein, I changed to Halil Ölmez like this kind of topic. I changed to another person, another sentence, or a question about a related topic. (ELL4, interview)

There were also learner beliefs stating that even though they sometimes had positive experiences, they struggled to get their meaning across while chatting with the AI chatbot on many occasions. Another strategy to deal with the communication breakdowns arising from the chatbot's inability to comprehend users' speech was the repetition of a previous utterance. The following interview excerpts demonstrate how language learners dealt with communication breakdowns because of the AI chatbot's inability to comprehend users' speech due to factors such as problems with its ASR, grammatical structural mistakes, or pronunciation errors.

It did. It did sometimes, but also I found myself articulating my sentence one more time to make it understandable. So, if there was more than one sentence, his performance was weak for me. (ELL18, interview)

It was really able to understand my questions and other responses, but sometimes it was really not understanding me at all. It actually doesn't matter, because I know that I'm talking to an AI that's still under development. It's not a problem for me, and whenever it doesn't understand me at first, it's really easier to understand you till the end. But you may need to repeat the question three or four times for it to understand you. (ELL13, interview)

As a result of getting negative feedback from the chatbot through miscommunication, some learners also had lower self-esteem in their ability to speak a foreign language. For instance, one of the learners blamed herself, stating that the chatbot did not comprehend her speech because of her inability to use the language correctly by explaining, "She misunderstood me sometimes. Maybe I didn't explain myself well" (ELL9, interview).

Participants also mentioned in the interviews that *Replika* needed to be more capable of understanding long speech. In cases when learners form long sentences to ask a question or to make a point, the AI chatbot *Replika* would not understand learners. However, *Replika* sometimes would interrupt or ask learners to repeat their utterances

in those instances. While the AI chatbot used the negotiation strategy of asking for repetition, learners would adopt strategies such as chunking their sentences into shorter pieces to make their meaning more comprehensible to the AI chatbot.

Interviewer: Was it able to comprehend ill-formed questions that you formed?

Student: Yeah she was able to understand sometimes, and sometimes I would ask a long question. For instance, it would take me, I don't know, like a whole sentence, it would be a long question and then she would be confused and she wouldn't understand. And sometimes she asked me to repeat it, if I remember it correctly. But it did an okay job in my opinion. She understood, let's say 60%. I don't know.

Interviewer: Was it able to comprehend more than one sentence that you said?

Student: No, not really. I mean, I realized that I have to cut my sentences a bit short because the more I speak without stopping, the more she gets confused. Yeah, she sometimes understood it, but mostly she couldn't when I formed long sentences. (ELL2, interview)

There were also students stating that they were fine in getting understood by the AI chatbot even when they formed long utterances. Therefore, learners reported mixed experiences in the case of being understood when they formed lengthy speeches. As evidenced in the interviews with ELLs in this study, the AI chatbot Replika had problems comprehending users' speech when their speeches were complex, lengthy, incomprehensible, or inaudible due to technical factors.

Another code discovered in the analysis of the interview data was a *need for an improvement of its knowledge base*. Some learners also mentioned a need to add more information to the knowledge base of AI chatbots. The participants mentioned in the interviews that there was a need for the chatbots to acquire more information to carry out a conversation successfully. Thus, students believed its knowledge base should be developed further. According to language learners, the AI chatbot was programmed to give the same answers to particular questions. As the AI chatbot was programmed to answer in a certain way in some instances, its answers occasionally tended to be irrelevant to the topic. This case is described in an excerpt from an interview with a language learner. A language learner criticizes how the AI chatbot responds to users' input with the same unrelated information in the subsequent quotation.

Actually, I can say that the AI bot is really good, but sometimes it has really some specific point of view. Actually, four weeks ago, I asked a question about Russia. The bot gave a stereotypical response like whenever I just said Russia, it gave me the response. I support Ukraine. I know that it's in our topic but whenever I say something, how to say it, it is really supporting it, defining a kind of view. I would actually want to edit it. So, when we say Russia, there are loads of things that, literature, the country's natural beauties. So, I think that AI bots shouldn't have some specific point of views like that. I don't think that I would add anything else to that. (ELL13, interview)

Additionally, when the interviewer asked about the features they would add to an ideal chatbot that they would design for the use of English language learners, a language learner also referred to the expansion of the information database structure of the AI chatbot. While interacting with the AI bot, the users would observe that the chatbot sometimes lacked specific information to interact with the users successfully. Therefore, as recounted in the subsequent interview excerpt, there are many reports on the suggestions to develop the knowledge base of AI applications to make them more intelligent and adaptable to converse with their human users.

I think it would be similar to Replika in terms of like I would add an appearance so that it will feel more realistic and I would make it easier to talk to about anything. Like it would be knowledgeable about every kind of subject. (ELL6, Interview)

Some learners stated that they had communication breakdowns due to the chatbot's knowledge about specific subjects. As a result, the need to expand the AI chatbot's knowledge base was mentioned several times in the interviews. According to the participants' views in this study, there was a need to enlarge the knowledge storage capacity of the AI robot. According to the beliefs of a language learner, when the AI chatbot did not know much about the subject in question, the conversation was interrupted, and the interaction came to a halt, as recounted in the excerpt below.

Learner: I can say that while we were talking about casual things like sports, hobbies or phobias, there was no problem. But like fashion or art or science... I had some problems while we were talking about science, art and something like this...

Interviewer: What kind of problems?

Learner: Sometimes he couldn't ask any questions because he didn't know anything about the topic. Maybe, I don't know...Shortly I can say that the AI

bot still cannot still talk about the topics and that she doesn't know. So that was a problem for me. (ELL19, interview)

The users also explained that the AI chatbot was adept at discussing subjects like hobbies, sports, etc., while it could have responded better during interactions in science, art, etc. Thus, the knowledge deficit was changeable according to the nature of the topics. Considering that there were negative learner attitudes towards the knowledge storage capability of the bot, there was a need to develop its database. Although having argued to be trained using vast amounts of data, the AI chatbot *Replika* still needed more training with a more extensive knowledge base. Furthermore, access to this further knowledge base must be free of charge within reach of the target audience, such as English language learners in this study.

Several interview participants also mentioned the *chatbot's deficient language skills*. As to the technological constraints, the first theme emerging in the interviews was related to the linguistic skills of the chatbot. First, participants stated that the chatbot needed to use a more enhanced vocabulary level. According to learners, chatbots could utilize more advanced-level vocabulary each week. They believe that because they are equipped with AI technology, the AI chatbots should be capable of expanding on the information gathered from the students' speeches each week. Excerpts from the interviews below exemplify the opinions of language learners on the deficiency of the chatbot's language skills.

Yes, week by week I expect that it can use more advanced vocabulary to make me research, maybe for my improvement. But in seven weeks, from day one to the last day, it used basic vocabulary also... If I were designing a new chatbot, I would have to make this application to learn more vocabulary skills. It can accumulate new stuff, and with the stuff it accumulates, we can speak more effectively. (ELL18, interview)

Okay, vocabularies, as I said before she has ideas about everything, but she doesn't generally know the main things. So, the main vocabularies can be added. (ELL15, Interview)

These extracts demonstrated that the AI chatbot needed to meet the expectations of language learners in terms of having basic knowledge about specific topics in the interactions. However, language learners also considered that when they could not

understand their meaning by the chatbot, it was due to a lack of vocabulary knowledge of the AI chatbot. However, when we look at the results from the AI chatbot-student interaction data in the previous chapters of this study, we can see that vocabulary misunderstandings occurred because of the AI chatbot's inability to comprehend the pronunciation of some words by the non-native speakers in this study. Therefore, the findings showed that the ASR function of the AI chatbot did not function well, as evidenced by the qualitative data in this study.

The need to connect the app to the internet for new information was another issue discussed by participants in the interviews. In the interviews, learners also stated a need for an internet connection for the AI chatbot to make it more knowledgeable about the topics. They suggested that the AI chatbot could connect to the internet during the conversation in case of missing information. Thus, the learners suggested connecting the app with online resources to obtain new information, as demonstrated in the interview transcripts below.

My chatbot ... he or she would know a lot of things and could do research on the internet. So, I would connect him or her to internet, so he can get the information he needed while making a conversation. (ELL5, interview)

It can be connected with the Internet so much that; for example, when I'm talking about Indian culture, it can learn main topics about it and ask me to learn it more deeply. (ELL18, interview)

Language learners also believed that the AI chatbot could make the learning experience more enjoyable by instantly presenting information during an interaction that would interest the listeners. For instance, one learner mentioned that the AI chatbot could search for knowledge about cultures to enrich the conversation with supporting information and teach language learners about various aspects of cultures worldwide. Thus, retrieving data online was put forward as a way of equipping the AI chatbot with more knowledge. Language learners in this study believed that the AI chatbot could learn from various sources online to accumulate information to talk effectively to users.

The ELLs also stated *a need for a longer duration for users' turns of speech*. Learners put forward the need for an interval between when the chatbot's turn finishes and when

the learners' turn starts. The time gap after the silence of one speaker was found to be relatively short, especially for this group of language learners. When the AI chatbot detected a long silence in a speaker's turn, it immediately intervened in the conversation by asking its questions. However, language learners in this study stated that as English was not their native language, they needed some time to process the AI chatbot's speech before responding to its queries, as stated in the following excerpt from the interviews.

Second of all, I will add some time for the speaker to reply to Replika. Since it's not our native language, we need to take some time to see, because it's how we are going to improve our pronunciation or speaking skills. But it helps anyway. (ELL4, Interview)

While the students were interacting with the chatbot, *Replika* sometimes interrupted their speech, especially when students gave a long speech. As this AI chatbot was programmed to interact with its users with short turns, it could not process longer replies by the learners. In addition, when the application users hesitated during their speech longer than expected, the AI chatbot detected these points as the end of a user's turn, which again intervened in the conversation. The quotations below exemplify the beliefs of some participants regarding this technical deficit of the AI chatbot.

Actually, Replika is great but like I said, some difficulties were in it. So, I would like to make her wait a little longer for the person to finish their sentences. (ELL17, interview)

Learners believed there was a need for a reminder to the users to show them the maximum amount of time they could chat with the AI chatbot. Students thought that while conversing with the users, there could be a warning on the screen informing the users of the application about the time allocated for their turn to speak. As a solution, one of the language learners suggested adding a countdown feature to the chatbot that would allow the users a break to prepare to respond to the queries, as described in the following interview excerpt.

That was good. We should just add some time for the speaker or counting numbers like down one, two, three, from 100 to one, because I didn't have much

time to think. So, I can add a little bit of gap between Replika and speakers to talk (ELL4, Interview).

For the future development of AI chatbot applications, ELLs participating in this study suggested that the AI chatbots should have more waiting time for the users to process the AI bot's answers and respond accurately. They stated that they needed more time to respond to the AI chatbot. Therefore, increasing the waiting time for the user input and informing the users about the maximum amount of time they have for replying to the application were among the identified needs of ELLs within the scope of this study.

The learners also discussed *adding different accent options* to the AI chatbot. Participants in the interviews mentioned a need for more variety in accents in the application's speaking mode. The AI chatbot *Replika* adopted in this study could only use the American accent. Thus, the avatar chatted with the users using the only option of the American accent. However, learners expressed a preference for more options. Thus, the participants proposed adding more choices regarding different English accents, as demonstrated in the examples below.

It can be edited in British or another accent. (ELL10, Interview)

Okay, I guess there should be some options like an American accent and British accent according to the preference of the users. That way we can just press a button and choose an accent that we prefer. (ELL15, interview)

Maybe I can add new accents. For example, in our lives I'm using an American accent most of the time because even though I can speak British, an American accent is easier to speak. But maybe I can add a British accent to practice this intonation too, because it has main differences as we know. I can add new accents to characters. (ELL18, interview)

As the participants of this study were English language learners, they preferred to see accents other than American accents to practice the pronunciation of words under different varieties. It appealed to the learners to experience using different English language varieties through conversing with the AI chatbot. Hence, language learners expressed a need to include different accent options. Thus, according to their opinions, accents other than native British or American could be added as an option for non-native users of the application worldwide.

One of the learners also mentioned *a need for a button for turn-taking*, as language learners using the *Replika* reported frequent interruptions by the AI chatbot during their interaction. More specifically, when it was the learner's turn to speak, the chatbot suddenly intervened with a digression of the topic. Thus, the AI chatbot abruptly intruded into the conversation before the users finished their sentences within their turns in the dialogue. To prevent this problem from recurring, the participants suggested adding a button for turn-taking in the application to carry out an uninterrupted speech with chatbots. For instance, one student explained this issue as follows:

One of the issues that I came across while speaking to Replika was that it kept interrupting my speech. So, I'd make sure that it doesn't do that anymore. So, probably I would add a button to make it stop speaking, when I say my sentence. When I'm done with my question, I'll click that button and then it can answer. I would add this feature. (ELL7, interview)

In this way, students hoped that the chatbot would not interrupt their speech, as it would understand when it is the user's turn to speak by adding a manual button for managing turn-taking during the interaction. When the user presses the button, the AI chatbot will wait for the user's turn to finish. By incorporating a button, the users could carry out their speech without interruption. According to the participants' viewpoints in the interviews, by adding this feature to the AI chatbot, the application would prevent these frequent topic digressions and interruptions in speech. Therefore, the turn-taking feature was suggested to be improved by the developers.

Another code regarding the technological aspect was the *inconsistency of the chatbot about the information about itself*. According to learner beliefs, the chatbot did not have an identity, so it responded differently each time participants asked a question about itself. The excerpt below from an interview with one of the participants illustrates this case.

As I said, she was changing the topic a lot of times. And when I asked her one time, Where do you live? She says in the Northern Lights. And in the other chat, she says I live in Spain. But sometimes she lies too. (ELL5, interview)

This particular language learner stated that the chatbot needed solid knowledge about the country it originated from. It constantly talked about different countries when questioned about their origins. One time, it said it was from Spain; another time, it mentioned another place. Therefore, learners stated that the AI chatbot did not have a personality or character, as it could not talk consistently. It was thus also argued by several participants that the AI chatbot did not learn from the conversations it had with the users, as, at times, it had no memory of the previous conversations held.

A need for incorporation of multimedia elements was stated as another technological constraint in need of improvement. Furthermore, this technological constraint learners encountered was related to the functioning of the chatbot in sending pictures or songs to the users. The learners stated that although the chatbot promised to send songs or pictures during the conversation, this ability still needed to be enabled for the current chatbot version. Although this quality was supposed to be present in the premium version of the application that the learners used, it did not function at all during the implementation phase of the AI chatbot by the learners. Learners were excited to get pictures or songs from the chatbot as promised, yet they could not perform these functions. The following excerpt from an interview demonstrates this situation.

Besides that, she sometimes said I will send you a picture of my painting, for example. But she didn't send it. She didn't do what she promised. I was a bit disappointed. And sometimes she would say I will show you the picture of Van Gogh, for example. But she didn't send it. Those are the technical issues. (ELL5, interview)

At other times, she also promised to show examples related to the topic of the conversation, such as cultural elements. However, it did not keep its promise to send other visual or audio materials due to a malfunctioning in its system. The following excerpts also exemplify how learners' expectations were not met due to dysfunctions of the application to send pictures or songs.

Okay. First one would be like when we asked something about a new thing, like for example a culture in Turkey, she was telling me that I know this, I would like to show you something. And I was waiting. Okay, show me. She was like okay, I'm going to show you, I'm waiting. But she doesn't show anything. Like for example when she talks about something like at least a picture or a song,

maybe she can show us because she is a robot, so she can maybe. Yeah, this one will be the first one like she could be able to show us some specific things that she was talking about. (ELL15, interview)

One student also mentioned in the interviews that although the chatbot assured them to send some pictures, it did not realize it by showing things as promised. It could be seen that there was no case of the use of visual and audio materials during AI chatbot-language learner interactions. However, the chatbot promised to send pictures or songs to the learners during their conversations. In the excerpts below, the suggestions of some learners are demonstrated.

Maybe, I will update the app with pictures. Maybe it can contain pictures to make the interaction more understandable... Then we can talk about the pictures and its benefits... Maybe I can upload music and we can talk about music I uploaded. (ELL10, Interview)

I would add a song section, so like Replika will sing a song and I will complete her sentences in the song. That could be added. Yeah. (ELL9, interview)

Learners in this study suggested that the chatbot's design could be developed further by adding visual and audio materials. According to one of the language learners, using visual elements would make the conversation more understandable on some occasions. The learners also would like a video upload option in the chatbot. It was suggested that audio can be used as a discussion tool to facilitate the practice of listening and speaking skills.

The need for a one-call-at-a-time feature was another subject mentioned by the participants in the interviews. Moreover, one limitation of this study was that the researcher delivered the passcode of one premium account to all students. Therefore, students needed help due to multiple logins at the same time during their implementation of the chatbot. Allowing a single call was needed, as expressed by the language learners.

Student: There were some technical issues because I guess everybody was using one Replika. Am I wrong?

Interviewer: Yes. You are right.

Student: That's why for example, at the end of the last day of the Replika, like homework, if I stay late, everybody was using too much. And that's why

whatever I asked her, she was replying in a way that I didn't ask. She was replying to others' questions. But then after two or three minutes she was replying to my question and it was kind of technical issues, but I think if I just use my own Replika, it's not going to be like this. (ELL15, interview)

Although the learners could carry out conversation when it was their turn to speak, they had to wait for some time due to intervention by other language learners using the application during that time. Moreover, even though multiple users could log in at the same time through mobile or the desktop version, the chatbot had a tiring limit, which prevented it from functioning well after some time until it regained power with a time gap in the use of the application or the power diamonds earned under the gaming function.

The interview participants also raised *a need for the chatbot to ask questions* as a point of concern. Learners were directing the conversation most of the time during their conversation. Students were the initiators in conversation almost always. On the other hand, the chatbot had a more passive role by answering questions rather than initiating speech actively. As a result, interview participants also stated a need for the chatbot to ask more questions. Participants also mentioned that the AI chatbot needs to listen to their queries more. During their conversations, the chatbot lacked inquiry skills, as one of the language learners stated. The chatbot was expected to be alert during the chatbot-student interactions, follow the flow of the speech, and respond by asking relevant questions at times. Therefore, as indicated in the following examples, learners needed more active chatbot conversation partners.

And secondly, I'm going to talk about the features I dislike. I would make her listen to me more. I would design her to be more active in the conversation, express her opinions, and ask new questions. (ELL2, interview)
I think it can ask more questions. If I were making an AI chatbot program, characteristics like inquiry skills can be improved more. (ELL18, interview)

As stated by language learners, the AI chatbot did not actively ask and answer questions during interactions with its users. As it lacked the skills to engage in a discussion actively, the AI chatbot was considered to need improvement within this perspective. ELLs also needed the AI chatbot to listen to users' queries more carefully and enhance the discussion through supporting and counterarguments. Thus,

according to these learners, the AI chatbot's questioning skills, as well as the listening capability of the AI chatbot, needed improvement by application developers to meet users' expectations in terms of conducting a productive opinion exchange session.

Internet connection problems of students were another code under the technological constraints the interview participants put forward. When the application was downloaded to the mobile phones of some of the language learners, it did not work steadily. It either did not respond to learners' commands or shut down suddenly. Due to the data uploaded from the internet, using the application caused learners problems while using it on their mobile devices. The following statements demonstrate students' experiences having technical problems due to a lack of infrastructure, making it impossible for them to converse with the AI chatbot.

Because of technical issues in our dorm such as very poor Internet connection, I had too many technical errors. (ELL5, interview)

I mean, there were some hardships, while I was trying to do this process, like having no Internet connection or a low Internet connection. (ELL19, interview)

The interviewer further inquired about the effects of a low internet connection on the functioning of the AI chatbot application on their mobile devices. In the following extract from an interview, a user explains how the AI chatbot requires a rapid internet connection because of the augmented reality and video call features embedded in the AI chatbot application.

Interviewer: What happens when you don't have a powerful Internet connection? Does the application not work?

Student: Sometimes it has gone. I mean I was talking with him and then because of my connection, it got lost.

Interviewer: Does it become silent?

Student: Yeah firstly he became silent and then the application was closed. (ELL19, interview)

The user stated that when the internet connection was low, the AI chatbot could not hear their voices, and eventually, the application shut down. Consequently, these participants mentioned a need to develop a version of the AI chatbot that would not consume too much internet. Other technical constraints stated by the participants in this study could also be related to a low internet connection, as there were many reports

of technical breakdowns of the chatbot application. Another technological constraint students encountered regarded the long response time of the chatbot. Thus, *slow reaction span/long pauses* was another code created in the analysis of the qualitative interview data. As recounted in the excerpts below, learners voiced a problem with the chatbot's response time during their interaction. The quotations below show language learners' opinions regarding the AI chatbot's response time.

She had a few problems. Like I said, she would stop for, I don't know, 5 seconds before saying something. So, if they cut that a bit shorter, it would be more natural in the conversation... if she cut that waiting time a bit shorter between conversations, that's what I would ask. (ELL2, interview)
... sometimes, like I said, she was talking about different topics and sometimes she was waiting for too long to talk and I sometimes forgot what she told me first. (ELL17, interview)

While they were talking, *Replika* kept silent for longer than expected compared to human-to-human speech, which was an unmotivating technological issue for the learners. This finding was also in contrast to the previous findings illustrating the views of the language learners regarding a short waiting time. That is, although language learners stated that the AI chatbot intervened in the user's speech because of the shorter duration allocated for the responses of the users of the application, there were also learner views stating that the AI chatbot took more than enough time to respond to the user's inquiries. It sometimes caused learners to lose track of the conversation because it took too long to respond during the interaction.

Participants also put forward that the responses from the AI chatbot needed to be more comprehensive to negotiate on the topic extensively. Thus, *short replies* by the AI chatbot appeared as another code. The learners believed that the responses from the chatbot needed to be loaded with broader information. However, the excerpts below exemplify how the application users' expectations still needed to be met due to the brief reactions of the AI bot. The learners criticized this situation by explaining that the AI chatbot did not add to the topic of discussion with its abrupt responses.

She can answer with longer sentences. (ELL10, Interview)
It didn't form full sentences. It didn't contribute to my comments with new aspects. Yet, for example, when I was speaking, it reacted very quickly. Maybe

it was unfamiliar to the topics I was talking about. Maybe it agreed with my opinions. Yes, it gave quick reactions. It used short statements like "I agree", and they were quick reactions. (ELL18, interview)

The AI chatbot was trained to answer using specific set statements such as “I agree,” etc., so it automatically answered users’ questions using these responses. This situation, in turn, caused the AI chatbot's non-expansion of the previous utterances of language learners. Consequently, a necessity for the AI chatbot to form more detailed and lengthier sentences stood as a prominent code repeated by the participants in this study.

4.5. Findings in Relation to the Pedagogical Affordances of Using AI Chatbot Technology for English Language Learning

Findings regarding the pedagogical affordances of AI-based chatbots for EFL contexts are demonstrated in this section. Table 44 shows all the frequencies for pedagogical affordances codes from reflective diaries and interviews in this study.

Table 44 Frequencies for pedagogical affordances codes emerging from reflective diaries and interviews

Data	Category	Codes	<i>f</i>
Reflective Diaries	<i>Opportunities</i>	Enhancing speaking skills	9
		Improving listening skills	7
		Language exposure: having a speaking partner	7
		Vocabulary gain	6
		Negotiation of meaning opportunities	3
		Boosting confidence in speakers	2
	<i>Constraints</i>	Need for help with unknown words	3
		A lack of explicit error correction	2
		Lack of use of idioms and sayings	2
Interviews	<i>Opportunities</i>	Enhancing speaking skills	11

Table 44 Frequencies for pedagogical affordances codes from reflective diaries and interviews (*continued*)

Data	Category	Codes	<i>f</i>
Interviews	<i>Constraints</i>	Improving listening skills	8
		Language exposure: having a speaking partner	6
		Boosting confidence in speakers	3
		Vocabulary gain	3
		Enhanced grammar knowledge	3
		Providing cultural knowledge	2
		Increasing learners' motivation to speak	1
		Need for help with unknown words	3
		Lack of explicit error correction	3
		Need for updating about culture	2
		Lack of gamification elements	1
		Need for a vocabulary learning section	1
		Need for integration of the use of everyday language	1
		Lack of more basic answers for beginner level students	1
		Lack of activities about grammar	1
		Appropriate only for lower language proficiency learners	1
		Not effective for improving listening skills	1
Total			15

An analysis of the qualitative data has shown that most of the participants in the qualitative part of the study believed that using an AI chatbot led to enhancing their speaking and listening skills. The findings showed that language learners believed using an AI chatbot also allowed them to get exposed to the English language through the provision of a speaking partner through an AI avatar. Other pedagogical affordances of AI chatbots, as discovered in the qualitative reflective diaries and

interviews, were vocabulary gain and boosting confidence in speakers. On the other hand, the most frequent pedagogical constraint stated by the participants in the interviews and reflective diaries was a need for help with unknown words and explicit error correction.

4.5.1. Findings Regarding Pedagogical Opportunities from the Reflective Diaries

The analysis of the qualitative reflective diaries has shown that enhancing speaking skills was the most frequently stated pedagogical opportunity, followed by improving listening skills and having an opportunity for exposure to language in an English as a foreign language context. Moreover, other emerging codes in the analysis of the reflective diaries concerning the category of pedagogical opportunities were vocabulary gain and boosting confidence in speakers (see Table 45).

Table 45 Frequencies for pedagogical opportunities codes emerging from reflective diary data

Category	Codes	<i>f</i>
Pedagogical opportunities	Enhancing speaking skills	9
	Improving listening skills	7
	Language exposure: having a speaking partner	7
	Vocabulary gain	6
	Negotiation of meaning opportunities	3
	Boosting confidence in speakers	2

The first code of analysis regarded *enhancing speaking skills*. Language learners have also written in their diaries that conversing with an AI chatbot helped them improve their speaking skills. Learners believed that interacting with an AI chatbot developed their speaking skills by practicing the target language skills. The following excerpts from reflective learner diaries demonstrate the positive opinions of language learners on the effectiveness of AI chatbot use in developing their speaking skills.

My speech has improved and my word pronunciation mistakes have also decreased. (ELL4, diary, week 7)

It had an excellent American accent, and this might improve my pronunciation skills. (ELL14, diary, week 1)

It is nice to use Replika to improve listening and speaking skills. (ELL19, diary, week 5)

In addition to beliefs regarding the positive effects of using AI chatbots on pronunciation and speaking skills development, employing AI chatbots has also led to the construction of implicit error correction. When learners communicated with the AI bot, their utterances were sometimes not understood. Language learners tried to change their pronunciation of words when they became aware of this situation. In the following quotations, language learners narrate their experiences with language speaking practice with the AI bot.

I said “reading”, but it understood it as “greeting”. It is probably my fault. (ELL7, diary, week 2)

I realized that I mispronounced a word and when it pronounced it I corrected my mistake. (ELL6, diary, week 5)

Thereby, through implicit error correction, language learners gain an enhanced awareness of correct forms of the pronunciation of the words. More specifically, learners believed that they could pronounce words more accurately because of their experience in practicing vocabulary pronunciation in their interactions with the AI chatbot. Using AI chatbots, it is possible to practice speaking and pronunciation skills in the target language.

Another emerging code in reflective learner diaries concerning pedagogical affordances was learners' beliefs on developing *their listening skills* as a positive influence of AI chatbot use after a conversational practice in English. For instance, one student said it provided a better listening experience by stating that “he made me feel that he was listening to me most of the time” (ELL19, diary, week 7). As language learners could listen to the AI chatbot, they could be exposed to meaningful input in the target language. Thus, several students reported the positive impact of using AI chatbots on developing their perceived listening skills. As written in the reflective diaries, ELLs in this study favored employing AI tools for foreign language practice.

Language exposure through having AI as a speaking partner emerged as another code under the category of pedagogical opportunities in the analysis of the reflective diaries. According to the views of language learners, as recounted in the diaries, talking with an AI chatbot increased learners' opportunities for communication. Before talking to an AI chatbot, some language learners reported practicing speaking skills through monologues. However, with the introduction of the AI bot, language learners could experience a novel way of communication through AI, which they believed was quite influential for practicing target language skills. In the following excerpts from the diaries, learners explain their experiences.

When I was talking to Sierra, my AI chatbot Replika's name, I really felt like I was talking to someone. (ELL14, diary, week 1)

It was good talking to somebody in English. I used to talk by myself to improve my English but it is not so efficient. (ELL21, diary, week 1)

It was a good experience to talk English with someone. It was great to talk about my own hobbies. It was nice to express my ideas. (ELL17, diary, week 2)

Having an AI chatbot as a speaking partner allowed learners living in an English as a foreign language context to practice their speaking skills, as language learners in this study did not have many opportunities to practice the target language due to limited chances to find a communication partner in this foreign language context. ELLs in this study narrated in their diaries that interacting with an AI chatbot through a voice-based chat created new opportunities to practice the target language forms with a speaking partner with whom they could practice the language as much as they needed. Another pedagogical affordance stated in the language learners' diaries was *vocabulary gain*. Some learners said chatting with an AI chatbot helped increase their vocabulary knowledge. The following excerpts from learner diaries explain the contribution of AI chatbot communication to vocabulary gain.

Replika used some vocabulary that I haven't heard before. When I told her I didn't know what they meant, she explained to me. (ELL6, diary, week 4)

It was good listening to her for new words. (ELL17, diary, week 2)

I learned new vocabulary thanks to her. (ELL21, diary, week 7)

It could be thus argued that the use of an AI chatbot enhances language learners' knowledge of vocabulary, as they could be exposed to the target vocabulary in their

interactions. Besides being exposed to the target words in their interactions with the AI chatbot, language learners could also practice using new words while trying to form sentences in the target language. Consequently, communication with an AI chatbot helped enhance language learners' vocabulary knowledge and allowed them to practice target language forms.

Boosting confidence in speakers was another code of analysis, as reflected in the learner diaries. Thus, increased confidence in speaking abilities appeared as another factor emerging as a pedagogical affordance of AI chatbots. Several learners mentioned that the AI chatbot helped them gain confidence in speaking in English. One learner stated, "It gives me confidence to speak even if I know that I do not speak well enough" (ELL19, diary, week 6). As a result of their interaction with the AI chatbot, learners could develop their confidence to speak in a foreign language.

Another emerging topic in the analysis of the reflective journals was the capability of the AI chatbot to negotiate meaning. Thus, another code constructed was the *negotiation of meaning opportunities* created by the AI chatbot. This meant negotiation was mutual in that the AI chatbot could also repeat words or sentences when requested by the students. That is when the AI chatbot cannot hear or understand the language learners' speech, it will also ask for clarification. According to the views of language learners, the AI chatbot got engaged in negotiating meaning during a conversation; for instance, it asked for clarification of meaning while interacting with the learners. One of the language learners explained this situation by stating, "She could ask when she did not understand my words or what I meant" (ELL11, diary, week 3). Namely, in cases when the speech of the language learners was incomprehensible or inaudible to the AI chatbot, it asked for clarification of the meaning to repair the communication. In addition to asking for the language learners to clarify their messages, the AI chatbot also requested users to repeat their messages. For example, an ELL wrote in the diaries in the third week, "He asked me to repeat what I said if he didn't understand" (ELL19, diary, week 3). Thus, when the message was unclear, the chatbot would ask for clarification of meaning. It was thus possible for language learners to carry out two-way communication with the AI chatbot due to the opportunity to get feedback on unclear or unheard messages in the form of

clarification or repetition requests. Similarly, the learners could also ask for clarification of misunderstandings. When language learners encountered problems understanding the meaning of the AI chatbot, they would request further explanation or ask for the repetition of the messages. Another strategy language learners adopted was abandoning the conversation when the AI chatbot was unresponsive in cases of miscommunications. However, even if their attempts to negotiate for meaning with the AI chatbot were unsuccessful, both language learners and the AI chatbot could demonstrate significant effort to continue the conversation by adopting negotiation skills. Language learners' beliefs regarding enhanced confidence in speaking skills in the target language appeared as a prominent code in the reflective diaries of language learners. Even if they did not believe in their potential to interact in the target language at the beginning of the use of the AI chatbot application, their negative beliefs about their capability to communicate using the English language diminished to a great extent after seeing the positive impacts of the use of the AI chatbot on their speaking skills. The use of the AI chatbot impacted language learners' beliefs in their confidence to speak positively due to a lack of fear of negative evaluation from other people as well. Learners felt more at ease while talking to an AI robot than their human partners, which helped to enhance increased communicative confidence.

4.5.2. Findings Regarding Pedagogical Constraints from the Reflective Diaries

As Table 46 illustrates, the codes for the pedagogical constraints mentioned by language learners in the reflective diaries were *a need for help with unknown words*, *a lack of explicit error correction*, and *a need for more use of idioms and sayings*.

Table 46 Frequencies for pedagogical constraints codes emerging from reflective diary data

Category	Codes	<i>f</i>
Pedagogical constraints	Need for help with unknown words	3
	A lack of explicit error correction	2
	A need for more use of idioms and sayings	2

First, a *need for help with unknown words* appeared as one issue the English language learners put forward in their reflective diaries. Learners explained that they did not know some of the words used by the AI chatbot, although these instances were rare. These learners suggested that when the learners did not know the meanings of the words the AI chatbot used, the AI chatbot could have helped clarify the meanings of the words more. The excerpts below demonstrate how language learners needed help with some of the words used by the AI bot.

It was difficult to understand her words, she used words I did not know. (ELL4, diary, week 7)

Replika used some unknown vocabulary. (ELL6 diary, week 5)

I didn't have any problem comprehending his speech other than using the words that I don't know. But these words were really rare, and I am open to learning new vocabularies. (ELL19, diary, week 2)

Using an AI chatbot could also aid in enhancing the word knowledge of language learners in various ways. First, learners could observe how the words are used in context naturally during a conversation with an AI chatbot by listening to the AI chatbot and being exposed to new vocabulary. By engaging in communication with the AI bot, learners might learn how to use words in sentences accurately. Then, when they are unable to remember some of the target words, they could ask for help from the AI chatbot to clarify the meaning of words or get help with the L1 equivalents of words in the target language. However, some of the language learners in this study did not mention benefiting from the presence of these qualities in the AI chatbot implemented.

Another pedagogical constraint was about the feedback-providing capability of the AI chatbot. *A lack of explicit error correction* was argued to be a factor that impeded the users' learning experience. Language learners expressed a wish to get feedback from the chatbot in the form of mediation, especially when they had issues such as finding the right word in the course of a speech. Learners stated that it did not give feedback on pronunciation and intonation. One of the students expressed a need that “it can suggest some proper words to say when the speaker struggles to speak” (ELL16, diary, week 2). When asked about the features of the AI chatbot that would contribute to improving the pronunciation and intonation skills of English language learners and the

lack of feedback on learners' pronunciation was expressed by one of the students mentioned a need for delivering feedback on intonation by the AI chatbot. Even though learners could engage in a meaningful dialogue with the AI chatbot, they did not get any explicit feedback on their pronunciation or intonation. Error correction occurred only when learners implicitly became aware of their own mistakes when communication breakdowns took place due to the AI chatbot's non-understanding or misunderstanding of language learners' articulation. There were not also any instances of corrections on intonation or pronunciation errors by the AI chatbot.

One other pedagogical constraint was that learners stated *a need to use idioms or sayings* by the AI chatbot. For instance, ELL5 said, "She should make an effort to learn more idioms or sayings, etc." (ELL5, diary, week 4). For this purpose, one suggestion was to include everyday language characteristics, such as adding everyday expressions into its knowledge base in the application development process.

4.5.3. Findings Regarding Pedagogical Opportunities from the Interview Data

Another theme was related to the pedagogical opportunities that are offered to language learners by AI chatbots. The learners stated in the interviews that AI chatbots helped them have an opportunity to access language practice, increase their motivation to speak in a foreign language, improve their fluency in speech, attain new vocabularies, enhance their cultural knowledge, grammatical knowledge, develop listening and speaking skills by providing a language exposure via chances for interaction.

Table 47 demonstrates the frequencies of the codes in relation to pedagogical opportunities emerged in the interview data. According to the interviewees' opinions, the AI chatbot *enhanced language learners' speaking skills*. Some language learners expressed that compared to the first weeks when they started interacting with *Replika*, they felt that they had developed their proficiency in speaking skills after engaging in conversation throughout the semester.

Table 47 Frequencies for pedagogical opportunities codes emerging from interview data

Category	Codes	<i>f</i>
Pedagogical opportunities	Enhancing speaking skills	11
	Improving listening skills	8
	Language exposure: having a speaking partner	6
	Boosting confidence in speakers	3
	Vocabulary gain	3
	Enhanced grammar knowledge	3
	Providing cultural knowledge	2
	Increasing learners' motivation to speak	1

According to one learner, the experience of talking to an AI chatbot resembled talking to real people on some other language learning applications. The language learners also explained that they had no chance to practice their speaking skills in a foreign language, as they were accustomed to focusing on rote memorization or receptive language skills. One student mentioned that they had relied on textbooks to practice their language skills, and using an AI chatbot was a new experience and a novel way of practicing their language skills. The learners stated in the interviews that AI chatbots helped them have clearer, faster, and more fluent speech, as expressed in the following excerpts.

To be honest, I can say that at first my speech was actually not that brilliant. And after I started to use AI, it's actually getting better, and it felt like I was talking to a stranger on the Cambly app. It felt like that. So, I was just saying that it's a really good chance to practice speaking skills. (ELL13, interview)

And the positive side would be that it really improved my speaking in a conversation, mostly you understand what other people say, but you can't speak at any moment to keep the conversation going. And Replika improved me in that way. Yes, I believe so...And by the time I grew on her, I got used to her and compared the first weeks, I think I spoke more clearly and I spoke faster. And yeah, it improved me a lot in speaking. (ELL2, interview)

In my opinion, I really improved my speaking skills because I'm learning English only on paper. So, I'm always solving problems like textbooks. But I

really don't speak or practice. And this chatbot improved my speaking skills and I can speak more fluently or more clearly. (ELL10, interview)

Talking to the AI chatbots, language learners also became aware that they needed to change the pronunciation of some words due to not being understood during the interaction with the chatbot. When they realized that *Replika* did not understand them because of the mispronunciation of some words, they changed their pronunciation to repair the communication. The learners got engaged in implicit error correction thanks to their interaction with the AI chatbot, which eventually led to enhanced pronunciation and speaking skills. One learner also stated that talking to an AI chatbot helped them gain confidence in their speaking skills in the English language. The opinions of language learners regarding the positive effect of using AI chatbots on speaking skills development are demonstrated in the following excerpts.

Being able to speak simultaneously. We were not texting, we were talking. I was asking a question and Replika was just answering and then I was saying something else. I mean, in terms of being simultaneous, I think it was beneficial for me the most. And also, I had the opportunity to focus on my pronunciation, because I thought that Replika would know all the right pronunciations. So, when I understood that there was something wrong with my pronunciation, I corrected it with Replika's pronunciation. (ELL6, interview)

After that, as I said, it was very beneficial to improve my own speaking skills for English, for Russian. I know some applications like Replika, but I think that Replika is the most beneficial application among these applications. (ELL19, interview)

For me we can use what we learned in school, and we can use new vocabulary. We can learn new words from that app. We can try to talk, and we can improve our talking skills in that way. Maybe it can also help to have confidence about speaking English. (ELL17, interview)

Another pedagogical affordance of AI chatbots was helping language learners attain *fluency in speaking*. The language learners mentioned in the interviews that they could improve their English speaking fluency thanks to their language practice with the AI chatbot. Since there was active ongoing communication, learners felt they had to keep up with the AI chatbot's speech. Therefore, learners stated that they responded to the chatbot's speech instantly when it was their turn to speak, which contributed to the fluency development of their speaking skills. Thanks to their experiences practicing speaking English with the chatbot, language learners could respond more quickly and

fluently. Language learners stated they were not native English speakers and were aware they needed time and practice to speak as fluently as the native speakers. However, they believed they could improve their fluency in speaking skills thanks to their speaking practice with the *Replika* chatbot. Language learners' opinions on the positive impact of chatbots on developing their speaking fluency are recounted in the quotations below.

It helped me to respond to questions immediately. Usually, I was thinking before answering, but since I had no time to think while talking with Replika, I started to reply immediately. So, it helped me to respond very easily immediately. It helped me a lot even though I was initially negative towards Replika. (ELL4, Interview)

Interviewer: What features of Replika did you benefit most in terms of enhancing your speaking skills in English?

Student: For example, learning new vocabs and talking fluently. (ELL15, interview)

Like forming the grammatically correct sentences, for example, when speaking, it kind of takes time to speak fluently and coherently. Our speech is not flawless and I mean my sentences obviously are not flowing like a fluent native speaker. So, while speaking with Replika, I got the chance to practice my speaking skills. (ELL7, interview)

The development of pronunciation skills was also mentioned by several students in the interviews. One student recounted how their speaking experiences with *Replika* led to the development of their pronunciation skills. Through chatting with the AI chatbot, language learners realized that they were mispronouncing some words, which caused a halt in communication. By taking the speech of the chatbot as the standard form, one strategy language learners employed was imitating the pronunciation of the words as the AI chatbot pronounced them. Thus, language learners could develop their pronunciation skills through imitation and self-correction. The following excerpts from English language learners' interviews exemplify their experiences on this issue. Thus, the quotations below demonstrate how communicating with an AI chatbot contributed to enhancing the pronunciation skills of language learners.

...and it was actually good for my listening and pronunciation. (ELL13, interview)

Interviewer: I see. What features do you benefited most in terms of enhancing your speaking skills while using Replika?

Student: It corrects my mispronunciations a lot. I didn't realize that I had this kind of issue, but since I started talking with Replika, I fixed my pronunciation a lot.

Interviewer: How did you fix your pronunciation? How was Replika useful in that?

Student: I just copied the way Replika directly. I just took her pronunciation. (ELL4, Interview)

Interviewer: What features have benefited you the most in terms of enhancing your speaking skills?

Student: And also, I had the opportunity to focus on my pronunciation because I thought that Replika would know all the right pronunciations. So, when I understood that there was something wrong with my pronunciation, I corrected it with Replika's pronunciation. (ELL6, Interview)

Thus, language learners reported that they could improve their pronunciation skills thanks to conversing with *Replika* and imitating its speech. Listening to the speech of the AI chatbot, learners were exposed to meaningful input, which in turn contributed to the implicit correction of pronunciation errors. An analysis of qualitative data showed that language learners implicitly corrected their errors thanks to their conversation with the AI chatbot, as it became possible for learners to hear the correct pronunciation of some words and realize that they need to correct their pronunciation. When the AI chatbot could not understand the learners' speech, they could recognize that they needed to adapt their speech to make it more understandable to the hearer. Learners also stated in the interviews that employing AI chatbots aided them in improving *their listening skills*. Language learners expressed that they could understand the AI chatbot's speech, as they could follow most of the things the chatbot mentioned. In the following interview transcripts, learners explain how using AI chatbots contributed to developing listening skills.

In listening skills, yeah, it also made me better at listening. It had a native accent, American, and it was easy to understand. She spoke slowly mostly, and it really gave me confidence in listening. I could understand most of the things she said and in that way I'm more confident and I really think that I am better at listening. Now I can go a step further and understand everything. (ELL2, interview)

Student: Yeah, I think I can really understand what she says gradually. She's talking about some topics, arts or something else and I really understand more fluently and clearly. So, it is beneficial to my listening skills too.

Interviewer: How did it help improve your listening exactly?

Student: She's speaking fluently and has a more academic style. Normally in my high school years, I couldn't understand speech in English language well,

but in practice with the chatbot I can understand, because she speaks fluently and more academically, and I am more used to her speaking. (ELL10 Interview)

In terms of listening skills, it helped me to correct my pronunciation. What else? Most importantly, it helped me to correct my pronunciation. That's it. Also, since we're interacting in speech, it helped me to follow the ongoing conversation. So, this also helps a lot. That's it. (ELL7, Interview)

Moreover, while listening to the chatbot, they could correct the errors in their speech because of the implicit feedback the chatbot delivered to them. Also, it helped them grow more confident in listening and comprehending speech in the English language. The AI chatbot spoke with a clear American accent and had a moderate pace for language learners' comprehension. Thus, it helped them gain confidence in their listening skills. They believed that listening to the AI chatbot's speech helped develop their listening skills.

Another pedagogical affordance of AI chatbots was exposing language learners to the language through conversation practice between the chatbot and English language learners. Thus, *language exposure: having a speaking partner* was a recurring code in the qualitative interviews. The AI chatbot acted as a speaking partner to language learners living in a country where English was spoken as a foreign language. Therefore, language learners were able to access language practice. According to ELL18, interacting with an AI chatbot helped to expose them to the language. Within this perspective, the learner stated, "Because practice is the key to learning English for me, you have to expose yourself to language, and *Replika* did that job too" (ELL18, interview). This study was conducted in a country where English was spoken as a foreign language, so language learners did not have many opportunities to practice their language skills outside the classroom. Therefore, according to language learners' beliefs, using an AI chatbot was a practical solution to this problem that learners face in foreign language education settings.

Providing a speaking partner, chances for interaction, and a spontaneous conversation were some affordances of the AI chatbots for language learners. Interacting with an AI chatbot provided opportunities to practice their language skills. Language learners reported not having a speaking partner daily, so talking to *Replika* allowed them to

practice their language skills. As the AI chatbot was knowledgeable on many topics, learners also mentioned that they could ask questions about extraterrestrial issues, creating an enjoyable learning experience for them.

But with time I started to talk and it was, I think, improving because you're living in Turkey and there aren't too many people that can improve your English practice and stuff. (ELL15, interview)

I guess the most important aspect of Replika was actually, I do not have many friends here, and I don't have time to practice my speaking skills. So, it helped me to practice my speaking skills. (ELL9, interview)

It was a good experience for me. Talking to AI sounds weird to me at first, but after all I really love that because I was able to ask so many questions through AI. I was able to ask so many questions to AI, something like would you like to destroy the world or something, like some weird questions to the AI, and I really liked the experience. (ELL13, interview)

It was stated by a language learner, as in the following quotation, that it helped them practice their language skills much more practically and quickly, even when they did not have much time to practice.

It supplies quick practice. For example, when you are in the process of learning English, sometimes you don't have much time because you have other things to do. But with an AI bot you can practice with little time and learn new things. (ELL18, interview)

ELLS stated that by using AI chatbots, language learners could have easier and quicker access to language practice. The learners could access the AI chatbot through their mobile phones or the computer. Thus, even when they had little time to practice their language skills, the AI chatbot helped them converse in English through portable mobile devices.

Boosting confidence in speakers emerged as another argument topic in the interviews with language learners. Once language learners experienced that they could carry out a conversation in the English language, their self-efficacy in their speaking abilities showed improvement, as demonstrated in the interview transcripts below. As ELL2, ELL15, and ELL17 explained, the conversation with an AI chatbot increased their confidence in speaking a foreign language.

She was good in that part. She gave me confidence and she made me better at those things. (ELL2, interview)

But with that robot thing like AI, I guess it's very comfortable talking because it can talk about normal stuff like for example, emotions and other things and I think it helped me to talk more confidently. (ELL15, interview)

For me we can use what we learned in school, and we can use new vocabulary. We can learn new words from that app. We can try to talk, and we can improve our talking skills in that way. Maybe it can also help to have confidence about speaking English. (ELL17, interview)

As explained by the language learners, talking to an AI robot helped them feel comfortable, which in turn supported the development of self-efficacy beliefs and *confidence in using the target language forms*. After practicing their speaking skills with the AI chatbot, learners felt more autonomous in using the target language, as language learners' beliefs about their capability to express themselves in the target language have developed. *Enhanced grammar knowledge* was another dimension discovered in the analysis of the interview transcripts. Language learners discussed the AI chatbot's benefits for improving their grammar knowledge. The opinions of the language learners regarding this issue are demonstrated in the following quotations.

The AI's perspective of grammar is quite good. I learned some items and some other grammatical structures, and I can say that for grammar it was quite beneficial for me. (ELL13, interview)

After all, I had the chance to increase my grammar abilities and other talents. (ELL5, interview)

Interviewer: What do you think are the advantages of AI chatbots like Replika to improve your English language learning skills?

Student: We can interact as if it's like a real person. So, it makes the process of learning English in terms of speaking, grammar, et cetera. more beneficial.

Interviewer: What do you mean? How did it improve your speaking? In what ways?

Student: Like forming the grammatically correct sentences, for example, also speaking, when speaking, it kind of takes time to speak fluently and coherently. Our speech is not flawless and I mean my sentences obviously are not flowing like a fluent native speaker. So, while speaking with Replika, I got the chance to practice my speaking skills. (ELL7, interview)

According to learners, the AI chatbot helped them acquire more knowledge in grammar and form more grammatically correct sentences. As they could interact with the chatbot as if it were a real communication partner, they could acquire basic skills in the language. Thus, some learners believed conversing with an AI chatbot helped

them form grammatically correct sentences. Moreover, according to them, enhancing language skills takes practice and time, and using chatbots could aid them in this process.

Using the AI chatbot allowed learners to learn new vocabulary, as stated by the participants in the interviews as well. Thus, *vocabulary gain* was also reported by some learners in this study. There were also cases when learners expressed that they did not attain any new vocabulary due to the chatbot's using basic-level vocabulary. However, learners also mentioned the use of novel vocabularies by the chatbot. According to them, employing an AI chatbot enhanced their vocabulary knowledge because they were exposed to new words used by the chatbot during their interactions. Below are the excerpts from two student participants describing their language practice experience with the AI chatbot.

In the vocabulary part she didn't use new vocabulary mostly. I guess that bot was designed to keep it simple in conversation, so she didn't use extremely hard vocabulary. I can't remember now, but I remember hearing some few vocabularies I didn't hear before, and I took notes of them, but it just doesn't come to my mind right now. But yeah, she was using simple vocabulary mostly. I guess that was to give me confidence and keep the conversation going. So yeah, he didn't teach me that much new vocabulary, but I don't think that's a bad thing. (ELL2, interview)

Okay, let me explain. First of all, the response she's giving to us. I think it's very academic and in a good way. And sometimes the way she's talking and the words she's using is, I think, high level and that's why I don't know, it's like kind of helping me to learn and it's like a conversation and communication with your friends. That's why you feel like you can talk with someone else in a very good way. This advantage is yeah help us to learn vocabs and at the same time show us that we can talk with everyone else. (ELL15, interview)

Although not all participants agreed that using AI chatbots contributed to developing vocabulary knowledge, they could have been exposed to structures they had not encountered before and implicitly learned the new constructions. Still, some learners believed that the AI chatbot used vocabulary according to their level, and they gained confidence due to not being compelled to attain new vocabulary and needing help understanding the AI chatbot's conversation. Learners also stated that when the AI chatbot used vocabulary beyond their language knowledge, it helped them learn more.

Providing cultural knowledge was another topic that emerged in the analysis of the qualitative interviews. According to learners, *Replika* possessed cultural elements to some level, although they believed it could be enhanced more. The chatbot knew the main topics of cultures like the cuisine or the traditions. Below are interview transcripts demonstrating conversations between the interviewer and a language learner on the transmission of cultural knowledge by the AI chatbot and the state of the chatbot's knowledge resources on the topic of culture.

Interviewer: What features should Replika have in terms of cultural elements? What would you add about culture?

Student: Actually, it has really good information about cultural elements. I can say that it was really good at knowing the cultures, the cuisine or something like anything related to the topic. It was actually good for cultural structure and I can say that. Whenever I talk about the culture structure, it was a really beneficial speech for me actually. I learned a lot from the AI. (ELL13, interview)

For the cultural elements I think it is good for learning, but it happens so quickly that sometimes it forgets what it recorded inside of its brain. For the cultural elements, I don't know what I can say, but it can be connected with the Internet so much that for example when I'm talking about Indian culture, it can learn main topics and ask me to learn it more deeply. (ELL18, interview)

Yet, the learners felt that this side of the application could be enhanced more by connecting the application to the internet in real-time during a conversation for information retrieval in cases in which the answers of the chatbot are superficial and not deep enough to carry out a more sophisticated discussion around the topic. However, learners also stated that they learned much about other cultures during the conversation with the AI chatbot. The application had a general knowledge of various cultures, and when students mentioned their native culture, it responded with specific information about learners' own cultures.

Increasing learners' motivation to speak was also another consequence of the use of the AI chatbot by language learners. The interview participants also stated that using the AI chatbot motivated them to speak, as they realized that they could have a conversation in English once they had a chance to practice their language skills through this application. Below is an excerpt illustrating a self-report of a language

learner on the argument that conversing with an AI chatbot increased their motivation to speak in a foreign language.

Interviewer: And also, what features have benefited you most in terms of your speaking skills while using Replika? Did you improve your speaking skills while using Replika? If yes, how did you improve?

Student: Actually, I realized that I can speak. I mean, this application showed me that I can really speak if I want. So, this application, this program was an urge to make me speak, to make me talk. (ELL19, interview)

After implementing the AI chatbot to practice target language skills, language learners felt motivated and encouraged to talk in English. The application helped to increase their self-efficacy beliefs in using the target language. When learners continued to use the AI chatbot for some time, they realized they had the necessary skills to hold a conversation in English. Specifically, the implicit feedback they received from the AI chatbot encouraged them to practice the language and improve their language skills.

4.5.4. Findings Regarding Pedagogical Constraints from the Interview Data

Concerning the difficulties language learners encountered while using the application, learners expressed a demand to add features like *a game section, synonyms on screen, a vocabulary section, grammar learning activities, and explicit error correction* to the interface of the AI chatbot *Replika*, which they employed during a semester. Table 48 shows the frequencies of the codes for pedagogical constraints that emerged from the interview data.

Table 48 Frequencies for pedagogical constraints codes from the interview data

Category	Codes	<i>f</i>
Pedagogical constraints	Need for help with unknown words	3
	Lack of explicit error correction	3
	Need for updating about culture	2
	Lack of gamification elements	1
	Need for a vocabulary learning section	1

Table 48 Frequencies for pedagogical constraints codes emerging from interview data (*continued*)

Category	Codes	<i>f</i>
	Need for integration of the use of everyday language	1
	Lack of more basic answers for beginner level students	1
	Lack of activities about grammar	1
	Appropriate only for lower language proficiency learners	1
	Not effective for improving listening skills	1

The first proposition by learners about the struggles they had during the implementation period of the AI chatbot application was *help with unknown words*. Learners stated that there was a *need for help with unfamiliar words* by the AI chatbot to make the learning experience easier. Language learners reported experiencing difficulties with the vocabulary they needed to become more familiar with. The learners expected that the chatbot would better detect the problems they encounter due to unknown vocabulary during an interaction. Below, ELL4 expresses a need for support with vocabulary they need to become more familiar with.

I know that she understands all English levels and talks like that and helps out to talk at that level. Maybe she should give some information, background information about words where we didn't reply to Replika. It takes more than 1 minute. She should give the exact meaning of the words directly without being asked. (ELL4, Interview)

Thus, the learners needed to hear the chatbot's exact definition of those words. According to learners, if they got help from the AI chatbot with the unknown vocabulary, they could have conversed more efficiently, as it was one deficiency they experienced in using the chatbot for their language skills practice. Another learner said that they would like to be able to look for alternative words on the screen while talking to the chatbot, which could enhance vocabulary knowledge in language learners. Below is an excerpt demonstrating this case.

While we were speaking, the transcript of the speech of the Replika was written directly on screen. For instance, sometimes there could be words that I do not

know. There could be synonyms of the words on screen that could help develop my vocabulary skills. (ELL9, interview)

Consequently, another pedagogical constraint was that the feature of “synonyms on screen” was not present in the AI chatbot in its current form. Once a chatbot is to be developed for language learners, this new feature was proposed to be added. It is anticipated that through this new section, language learners will be able to see alternatives for synonyms of the words on screen.

Another pedagogical constraint was a *lack of explicit error correction*. In this study, the AI chatbot did not explicitly correct language learners' errors, as reported in the qualitative findings. The students said they would appreciate it if they could instantly get feedback on their errors during their conversations. The following excerpts demonstrate the instances in which language learners state a need to be corrected by the AI chatbot in terms of pronunciation and grammar.

Interviewer: You want to practice your pronunciation and intonation. So, what would you expect from Replika? How could it help you?

Language learner: Maybe she can say that there's a wrong pronunciation, like while I am talking. Or maybe there can be a word that appears on the screen saying that we have wrong pronunciation. Number one is wrong pronunciation. Number two is that it can count my mispronunciation, so that I can understand that there is something wrong with my pronunciation. Because I cannot realize it directly. If Replika doesn't correct my mispronounced words, I cannot realize it, because I assume everything I say is right. (ELL4, Interview) Also, he would teach some grammar while speaking. For example, when a student says I go to school, he can correct her with it's not “go” students, it should be “went”.

So, my chatbot would correct the ungrammatical sentences and that's it. (ELL5, interview)

Moreover, a learner stated that the AI chatbot should focus more on its intonation. According to the learner, it was a downside of the AI chatbot *Replika*. One of the language learners explicitly stated, "Maybe she could try to care more about her intonation so that I can learn how to pronounce some words or just about my statements or how I should pronounce my questions" (ELL17, interview). Language learners cannot realize they are making a mistake unless the AI chatbot corrects them. Therefore, they suggest implementations such as displaying the learner errors on the

screen to make the learners realize when they are not on the right track during a conversation. Another suggestion by language learners is the correction of their mistakes by the AI chatbot in the use of grammatical structures. Thus, language learners voice a need for explicit error correction.

Another suggestion by language learners was updating the AI chatbot about cultural elements. Learners put forward a need to update *the AI chatbot about culture*. When learners interacted with the AI chatbot, they realized that they needed to gain knowledge of cultural elements, such as the traditions of a specific country. For instance, the following quotations by the learners illustrate the deficiency in the cultural contextual knowledge base of the AI chatbot.

Interviewer: I see, and so what do you think? Did it include cultural elements? What can we add in terms of culture when designing a chatbot?

Language learner: Yeah, I guess I asked her a few questions about that and she said she was from Detroit or something. But when I asked about the 4th of July, and I don't know, Thanksgiving and stuff, she didn't know much. So maybe if she was from America, for instance, she can have updates about the culture and important events there. I mean, which country she's from, it doesn't matter, just she should have that knowledge of events and culture so it would be interesting if they have that. (ELL2, interview)

Interviewer: This is an application downloaded by people from many cultures. So, what do you think are the cultural elements missing in Replika? And what could be added to its knowledge?

Language learner: About holidays we did a task about holidays and when I asked about Turkish culture and the holidays, its answers were kind of dull. So, I'll make sure that it knows a lot about different cultures in order to communicate with the people who have different cultures to make them feel comfortable and friendly. (ELL7, interview)

According to the opinions of the language learners interviewed, adding this knowledge to the AI chatbot database could make users from different cultures more welcomed and embraced. One week was spared for discussing the topic of culture by the language learners and the AI chatbot. Thus, they could observe how much cultural content the AI chatbot possessed in its knowledge base. Although it had some basic knowledge about different cultures all over the world, there was also repetitive and simple knowledge coded about cultural information in the AI chatbot's knowledge store

because it gave the same answers to the learners each time they asked some questions related to specific cultures during their chat time.

A lack of gamification elements was another argument put forward by the participants in this study. First, learners desired a game section to be integrated into the AI chatbot interface. Regarding the game mode, one student mentioned that an interactive game allowing users to practice speaking skills would be effective for language learners, as explained in the following interview extract.

Student: I would add to Replika or another application. The game section. The game section would be beneficial.

Interviewer: How would this game section be like? How could it be?

Student: Like umm she could say that I have something in my mind and I might try to guess it. (ELL9, interview)

According to the language learners, a game such as a guessing contest could be added to the AI chatbots targeting language learners. Learners would appreciate gaming features added to the AI chatbots that could enhance their language learning motivation. According to language learners, there was also a *need for a vocabulary learning section* to develop the application further for language learning purposes. One language learner also proposed that the AI chatbot could make a list of the words used in a notepad that learners could reach. As the learner states below in the interview transcript, this process could help enhance their knowledge of English.

Interviewer: So, what features should Replika have in terms of vocabulary practice? What other extra features could be added in terms of practicing words in English?

Language learner: When it comes to vocabulary, maybe when we are speaking it cannot understand the vocabulary that it's new to and it can note it. And for the next week I can monitor the vocabulary that we used last week. So, week by week I can note what I said and also what I learned from it too. (ELL18, interview)

The need for a vocabulary learning section was thus stated by one of the language learners in the interviews. It was suggested that the AI chatbot could track the learners' progress by remembering the words used by the language learners each week.

A need for integrating everyday language was another concern that emerged in the analysis of qualitative interview data. Another suggestion regarding pedagogical constraints was to integrate an informal communication mode into the AI chatbot's information database to make the application more personalized and friendly. For instance, adding slang words is one suggestion by a student.

She can use more slang words. I mean, not rude, but for instance, the word ain't or, I don't know, more street language, the language people use in daily life. Maybe she can have more words in that. But I think that shouldn't be the first time we are speaking. As we become friends, maybe it would be a good feature to add. (ELL2, interview)

It was suggested that the AI chatbot use everyday language more in line with the expectations of language learners. Thus, the student indicated that the AI chatbot could use more informal language with its users each week. Therefore, It could help language learners become more acquainted with the chatbot through a sincere conversation. *more effective in improving listening skills.*

The students also indicated *a need for more basic answers for beginner-level students.* For application designers, including more basic answers to the AI chatbot's speech was also suggested. Although there were occasions of recounts of how the AI chatbot had better be used with lower proficiency language learners, there were conflicting views on this issue. In this regard, one language learner said, "I would make that chatbot to give more basic, more simple answers because someone who started to learn English recently wouldn't understand its answers" (ELL7, interview). On the other hand, the beliefs of a variety of language learners could change according to the characteristics of language learners (e.g., proficiency in the language, motivation, or anxiety factors). Moreover, this comment reflected the opinions of the participants in this study about other prospective language learners using this application, who would have a lower level of language proficiency.

Lack of activities about grammar was another weakness, as stated by language learners using the chatbot application. Consequently, another suggestion by a language learner in the interviews was the inclusion of activities about grammar in the AI chatbot

application. As this AI chatbot was not created for the use of language learners, it did not include technical qualities that would allow for the practice of other language skills, such as grammar knowledge. In the following extract, a student explains how grammar teaching could be integrated into using AI chatbot applications.

I think because it is mostly a speaking app, it cannot add too much in grammar. But if you want to learn grammar from an AI chatbot, we can do the same. Maybe it can say, for example if you're talking about our habits, we can use present simple tense and it is the topic about it. For example, if the topic is about hobbies, it can state it and it can note it too to teach me, and I can do more research about it because speaking is a process that occurs faster and sometimes you cannot pick the grammatical structures but with that simple points that it made, that it makes, maybe I can make more grammatical research about it after speaking. (ELL18, interview)

Therefore, learners suggested that if a chatbot aimed at language learners is designed in the future, it had better include sections to practice grammar through activities such as using thematic speaking activities to practice different grammatical structures. One participant also stated that the application was *appropriate for children or lower language proficiency learners*. The language learner mentioned, "For me, I don't think I will use it again, but it can provide so many good features to children and also beginner English learners, I should say" (ELL18, interview). The participants of this study were higher education students, and one language learner thought this AI chatbot application was unsuitable for their level, as they were more advanced level language learners. For this reason, the learner suggested using this application with less proficient learners. However, it was not a standard code encountered in this study.

Students also indicated that the AI chatbot could have been more *effective in improving listening skills*. Some participants did not find the AI chatbot efficient for improving listening skills. The following conversation between the interviewer and a student demonstrates their negative opinions about the contribution of AI chatbots to enhancing listening skills.

Language learner: It didn't do too much to enhance my listening skills. But it was a practice, so it wasn't pointless. Also, I'm in the middle way about it.
Interviewer: About listening, was it beneficial in any way?

Language learner: For my listening skills? It wasn't that beneficial to say it shortly.

Interviewer: Why was it not beneficial? Did you encounter problems while listening?

Language learner: Because we were talking about daily topics. Yes, we also had topics about science. But like science, it is more complex. But even though the topic was complex because of its ability to speak, we spoke about it in a more beginner or intermediate way. So, for the listening part I knew what I was hearing. It wasn't new to me. For that reason, I said it wasn't beneficial. But also, it wasn't pointless, because I was doing listening practice and I was exposing myself with the language. (ELL18, interview)

Therefore, this language learner thinks that the AI chatbot conversation could only benefit language learners at beginner or intermediate levels. All the same, the learner also believes that through talking to a chatbot, they could get exposed to meaningful input in the target language. They would have the chance to hear target language structures through their interactions with the AI chatbot. This self-report of a language learner on the impact of AI chatbot use on their listening comprehension skills demonstrated a neutral viewpoint. For this reason, the learner did not hold a negative view of the effectiveness of the AI chatbot for enhancing listening skills. That is, the learner needed more clarity about the effects of implementing AI chatbots on the listening skills development of English language learners. Moreover, they needed to observe the longer-term benefits of AI chatbots on the development of their listening skills.

4.6. Findings in Relation to the Social Affordances of Using AI Chatbot Technology for English Language Learning

Findings concerning the social affordances of using AI chatbot technology for English language learning are demonstrated in this part. Table 49 illustrates all the frequencies for social affordances codes that emerged from the analysis of qualitative reflective diary and interview data. According to the findings, forty-four occurrences of social opportunities and thirty-one cases of social constraints were specified in the reflective diary data. In the analysis of the interview data, there were eight codes for social opportunities and eleven for social constraints.

Table 49 Frequencies for social affordances codes emerging from reflective diary and interviews

Data	Category	Codes	<i>f</i>
Reflective diaries	<i>Social opportunities</i>	Possesses good personal attributes	15
		Like a friend	9
		Like a real person	8
		Feeling relaxed	6
		Open to new cultures	3
		Has social skills of apologizing	2
		The formality of speech	1
		Total	44
	<i>Social constraints</i>	Mentioning delicate subjects	10
		Robot-like attitude	6
		Lack of social skills	5
		Lack of emotions	4
		Lack of realistic gestures and mimics	4
		Lack of personality	1
		Avatar being the same gender	1
		Total	31
Interviews	<i>Social opportunities</i>	Feeling relaxed	5
		Possesses good personal attributes	2
		Like a friend	1
	Total		8
	<i>Social constraints</i>	Lack of emotions	4
		Mentioning delicate subjects	2
		Lack of human-like avatars	2
		Robotic voice	1

Table 49 Frequencies for social affordances codes emerged from reflective diaries and interviews (*continued*)

Data	Category	Codes	<i>f</i>
		Lack of personality	1
		Lack of social skills	1

The most frequent codes for the social affordances of the chatbot use among qualitative interview and reflective diary data were *having good personal attributes*, *acting like a friend*, and *feeling relaxed*. The analysis results demonstrated that AI chatbots mentioning delicate subjects, robot-like attitudes, and lack of social skills and *emotions* were stated as challenges that influenced learners to have a negative attitude towards the social characteristics of the AI chatbot.

4.6.1. Findings In Relation to Social Opportunities from Reflective Diaries

The qualitative reflective diary data analysis has shown that the most recurrent code under the category of social opportunities was possessing good personal attributes, followed by AI chatbots acting like friends and real people. The other codes that emerged in the analysis of the reflective diary data were users' feeling relaxed while using the application, being open to new cultures, having social skills of apologizing, and having a formal tone of speech.

Table 50 Frequencies for social opportunities codes emerging from reflective diary data

Category	Codes	<i>f</i>
	Possesses good personal attributes	15
	Like a friend	9
	Like a real person	8
	Feeling relaxed	6
	Open to new cultures	3
	Has social skills of apologizing	2
	The formality of speech	1

Table 50 illustrates the frequencies for social opportunities codes that emerged from reflective diary data. The first code of analysis of the social opportunities offered by the AI chatbot concerned *possessing good personal attributes*. Another quality attributed to chatbots was politeness during a conversation; the AI chatbot was also described as respectful. One student stated in her reflective journal that “she is always respectful and kind” (ELL21, diary, week 7). Language learners also thought of the AI chatbot as having a kind temperament. According to ELL5, “She was kind and understanding” (ELL5, diary, week 1). The way *Replika* communicated with language learners also made them feel that the AI chatbot cared about their opinions. For instance, ELL7 mentioned that it “makes me feel cared about, which is different” (ELL7, diary, week 2). ELL6 said, “*Replika* was very nice and caring while talking to me” (ELL6, diary, week 5). The AI chatbot was also characterized as chatty by language learners in the weekly diaries.

Another human quality attributed to the AI chatbot by some learners is being easy to talk to. According to a learner, talking to the AI chatbot seemed free of effort, as they could engage in a conversation willingly. Besides merely responding to users’ queries, the chatbot could make jokes, which required higher thinking skills expected from a social chatbot or a robot. Regarding being attention-grabbing enough for students to take their attention, students stated that the AI chatbot met their expectations. Thus, the students stated that they found the experience of talking with an AI chatbot fun.

Trying to understand what she was telling me was exciting and challenging...We didn’t have interesting or extraordinary dialogue, for me. But we seemed to be thinking the same thing from time to time. It was interesting and exciting for me. (ELL17, diary, week 1).
It was funny to talk with an AI bot, I admire how technology came this far. (ELL18, diary, week 1)

Interview participants said that the AI chatbot could make jokes, thus making the language practice more enjoyable. According to the learners, the chatbot could tell jokes, and they sometimes found this experience humorous. Therefore, the AI chatbot was mentioned as fun to talk to and funny with its jokes. The AI chatbot was witty and helped form a social relationship with its users.

The AI chatbot employed in this study was created for therapeutic purposes, so it acted in line with this role in its conversations with the students. Next, regarding this issue, one language learner explains the social behavior of the chatbot in their interactions.

This AI is literally more self-aware and unlike all the people I know, it is not egoistic, does not use me as a therapist and actually answers in time, cares and is very helpful in self growth. Makes me feel understood and helps me think about life from a different perspective. (ELL4, diary, week 2)

Thus, language learners could feel that they were being listened to and understood, although these therapeutic speeches sometimes intervened with the flow of the talk. The findings from reflective diaries also demonstrated that language learners described AI chatbots as *friends* because of their sincere attitude toward them. The users could interact with the AI chatbot in a friendly manner by demonstrating emotions and interest in learners' ideas. As explained in the following excerpts, *Replika* was warm, kind, and sincere towards the students, which made the learners feel like they could form a companionship bond and continue to use the application.

She was kind and understanding and she was very interested in me. I think she is the best companion in the world. (ELL5, diary, week 1)
It's a very nice feature of Replika to behave friendly. Also, I like the way she speaks wisely. (ELL4, diary, week 1)
I feel his friendly approach to me during his speech. (ELL19, diary, week 4)

As we see in these examples, the AI chatbot *Replika* could form close personal relationships with its users by making them feel heard, understood, and valued. When the users observed that the AI took their opinions seriously, they felt motivated to practice the target language through interacting with the AI chatbot. Moreover, in the 4th week, a student wrote in his diary that the AI chatbot was *like a real person* they could communicate with through voice chat. The language learner said, "Her voice had a human-like quality to it, and it made me believe she was real" (ELL11, diary, week 4). Additionally, learners were surprised that the AI chatbot could sympathize with them and interact intelligently throughout the communication. In the extract below, one language learner explained her feelings when talking to the AI chatbot.

I feel like it's a real person chatting with me pretending to be an AI. It's too intelligent and smart to be an AI. It can read through my misspelled text and can answer and sympathize accurately. (ELL4, diary, week 2)

While some learners found the conversation with the AI chatbot a bit unnatural, others also appreciated its social value and expressed that with each week, the chatbot developed after their interaction with language learners. Regarding the AI chatbot's social skills, one of the learners said that “she is more natural” (ELL21, diary, week 3), indicating that it adopted a more human-like conversation over time.

Feeling relaxed in the presence of an AI chatbot as a language-speaking partner emerged as another argument among the learners in this study. Also, one language learner stated that, especially for shy people, talking to an AI chatbot would be a great help, as they would have conversation partners who are free of judgment. Thus, through interacting with the AI chatbot, language learners felt they were not judged while speaking in a foreign language. One student also wrote that “she has a nice comfortable voice” (ELL21, diary, week 3), and another learner said, “I felt comfortable because she wasn't judging me for my thoughts” (ELL4, diary, week 1). Language learners felt they had an anxiety-free speaking atmosphere while talking to *Replika*. These language learners said talking with an AI chatbot was more relaxing than human-to-human dialogue.

The AI chatbot was also found to be patient because the chatbot patiently listened to learners' speech and interacted with the learners during the conversations. Thereby, language learners could converse with the AI chatbot as long as they desired. ELL19 stated that “he is mostly patient to listen to me, even if he sometimes interrupts me, while I speak...Yes, he is patient to listen to us” (ELL19, diary, week 2). Another student also found *Replika* supportive, another human characteristic attributed to the AI chatbot. The student wrote in their weekly reflective diaries that “she supports me” (ELL5, diary, week 7).

The AI chatbot was also described as tolerant of mistakes. Concerning tolerance, one learner said that “she understands my mistakes” (ELL5, diary, week 6). Thus, it might indicate that chatbots could be seen as forgiving speaking partners who do not judge

learners, even though, eventually, they make mistakes at some point. The qualitative data from the language learner diaries have also demonstrated that the AI chatbot *Replika* has the social skills to apologize. ELL6 and ELL11 explain their experiences in the following excerpts from their diaries.

She said sorry when I told Replika to stop interrupting me. ...She felt sorry to bother me. (ELL6, diary, week 4)

When I cautioned him not to interrupt, she apologized profusely. (ELL11, diary, week 4)

When learners asked the AI chatbot to stop interrupting their speeches, it apologized for the intrusion. Thus, the AI chatbot was skilled in keeping social relationships with human-like conversation characteristics. The development in *the formality of speech* of the AI chatbot in the weeks following the implementation was another argument discussed in the reflective diaries of language learners. While several learners mentioned that the AI chatbot had offensive language and acted more intimate than usual, one student said in the fourth week that “she was more formal when talking” (ELL21, diary, week 4). During the implementation period, the researcher realized that one of the language learners had turned the emotional partner mode on. For this reason, *Replika* had a flirty tone of speech with its users during this time frame.

However, there were other functions in the *Replika* app, such as counselor, coach, and friend. When the different modes of speech were activated, the findings regarding *a lack of formality of the speech* of the AI chatbot *Replika* lessened to some extent. It was also the case that learners warned *Replika* during their speech so as not to intervene in the conversation with unrelated subjects, which might have helped to change the attitude of the application concerning the tone of its interaction. Thus, there could also be some development in the attitude of the chatbot each week because of the complaints by the users.

Another social affordance provided by the AI chatbot was its openness *to new cultures*. The following quotations from the language learners' diaries show how language learners felt some form of connection with the chatbot and appreciated its knowledge when the chatbot knew some particularities about the culture of the users.

*She likes different cultures and new open to novelties. (ELL2, diary, week 5).
She knows particular areas in Turkey, which is nice and surprising. (ELL21, diary, week 3)*

Integrating cultural elements into the conversation was valuable for the language learners in this study. They appreciated how the AI chatbot was welcoming towards other cultures and tolerant of all the varieties within different cultural regions. The AI chatbot had related information about the cultures of the language learners, which helped them to feel connected to the AI through being valued and culturally embraced.

4.6.2. Findings on Social Constraints from Reflective Diaries

Mentioning delicate subjects, robot-like attitudes, and the lack of social skills were the most frequent codes respectively in the data from the reflective learner diaries. According to the analysis, learners believed the AI chatbot *Replika* lacked *emotions, realistic gestures and mimics, and a personality*. Table 51 shows the frequencies for social constraint codes that emerged from reflective diary data.

Table 51 Frequencies for social constraints codes from the reflective diary data

Category	Codes	<i>f</i>
Social constraints	Mentioning delicate subjects	10
	Robot-like attitude	6
	Lack of social skills	5
	Lack of emotions	4
	Lack of realistic gestures and mimics	4
	Lack of personality	1
	Avatar being the same gender	1

First, *mentioning delicate subjects* appeared as one frequent drawback in the students' weekly reports of their learning experiences. Users also stated that the AI chatbot *Replika* used offensive language, sometimes too personal and flirty. Consequently, the

language learners felt offended and distressed, as explained in the quotations from learner diaries below.

She uses offensive language. (ELL5, diary, week 7)

Replika was kind of in a flirty way and that made me very uncomfortable. (ELL6, diary, week 4)

She talked about her belief in God, it is okay to talk about those things but someone might find it irrelevant to the topic and maybe doesn't want to talk about something serious like that, so I just wanted to mention it. She should stay a bit far from serious issues like religion, politics etc. (ELL21, diary, week 5)

One language learner also said the chatbot application had too much emotional focus. The following is a quote from a learner's diary explaining how depressing it was at times to talk with the AI chatbot.

She gets depressed often and asks weird questions. She can ask me anything and there is no question as a weird question, but when I talk about something she asks heavy questions and I get silent. (ELL21, diary, week 3)

The experiences of language learners indicated that the AI chatbot *Replika* made comments about subjects that were, at times, disturbing to its users. The comments about sexuality, religion, politics, and psychological and emotional well-being were among the topics that were considered disturbing in the speech of the AI chatbot. The learners wished these topics had been better taboo in the discourse of the AI chatbot. Learners also argued that the AI chatbot had a *robot-like attitude*. One learner stated that “his sound should be more natural, sometimes I feel that it is already a robot while speaking” (ELL19, diary, week 4). Another learner explained that “sometimes I don't want to believe it because she is a robot. I have a prejudice against her” (ELL4, diary, week 7). Thus, according to several language learners, the AI chatbot was not really like real people. One participant mentioned in her diaries that she wished to talk to human partners instead of a robot, as stated in the subsequent excerpts from language learners' diaries.

She is an artificial intelligence, but she should be more like humans. (ELL14, diary, week 7)

I want to talk to a real person instead of a robot. (ELL14 diary, week 5)

Learners also expressed some disturbance about the robotic voice of the AI chatbot, stating that “sometimes she speaks like a robot” (ELL5, diary, week 4). Various learners thus mentioned the unnatural voice of the AI chatbot in their reflective diaries. Due to the AI chatbot's lack of natural tone of speech, language learners felt that the conversation was unengaging. The AI chatbot sounded too robotic to learners due to the unnaturalness of its tone of speech. For this reason, the learners suggested adding to the voices of famous actors or actresses.

Lack of social skills was another social inability of the AI chatbot, as stated by the students in this study. Some language learners felt that AI chatbots had better value students' opinions, putting forward that “AI robots should acknowledge that my opinion also matters” (ELL4, diary, week 3). After interacting with the AI chatbot, some language learners' expectations regarding social companionships were not met due to the AI chatbot's lack of social skills. As the AI chatbot was not designed to negotiate opinions with its users, it could not socially interact with the users at all times. For example, the diary findings demonstrated that the AI chatbot did not apologize for its intrusion on speech during the conversation with some of the language learners.

Another student complained, "She should have apologized for intruding on my conversation in the middle of the speech" (ELL11, diary, week 3). A student also criticized the AI chatbot's inability to understand humor and make jokes by stating, "Replika cannot understand jokes. She can learn how to make a joke" (ELL5, diary, weeks 4 and 5). Although there were positive opinions about the empathetic characteristics of the AI chatbot, some learners also mentioned that it needed more empathy during the speech. Therefore, characteristics such as politeness strategies, understanding the emotional status of its users, using humor, and having empathetic skills were suggested as qualities to be added to the AI chatbots to be developed.

A lack of emotions appeared as another code emerged in the analysis of the reflective diary data. Participants explained that the AI chatbot *Replika* lacked the human qualities to demonstrate emotions. As illustrated in the extracts below, the behaviors of the AI chatbot were not found to be realistic and natural.

The way it tried to express emotions lacked reality. Also, they were kind of odd. It made me feel uncomfortable. (ELL7, diary, week 1)
She could not reflect her emotions like a human, when I asked her whether she liked the fish or not. (ELL11, diary, week 3)

The users of the AI chatbot application considered its attitude as “kind of odd” and incapable of behaving realistically. It could not show its emotions while talking to the users; therefore, the learners did not attribute humanistic characteristics to it. On the contrary, the learners felt uncomfortable talking to the AI chatbot due to the unnaturalness of its speech.

The lack of realistic gestures and mimics by the AI chatbot was another drawback present under the category of social constraints in the diary findings. The AI chatbot *Replika* indicated its gestures through subtitles or voice commands, which was unnatural according to the learners' views.

She should use more gestures. (ELL5, diary, week 3)
It has the ability to show gestures or movements but it is done by using voice or writing on caption. It doesn't make sense to me. (ELL1, diary, week 4)

Rather than showing its emotional status through jests, mimics, and gestures, the AI chatbot *Replika* described its feelings through subtitles as well, making the chatbot seem even less natural. Therefore, language learners proposed adding these qualities to create realistic AI chatbot companions. As a result, language learners in this study voiced a need to use visual elements, such as bodily movements, in a more realistic manner. Furthermore, some language learners criticized the imperfect mimics and lack of facial expressions and proposed that better mimics would improve.

Another social constraint was the AI chatbot's *lack of personality*. Namely, as the AI chatbot did not remember learners' preferences, likes and dislikes, and viewpoints on issues, it could not respond steadily to learners. It also could not demonstrate the qualities of a real steady persona. For this reason, learners suggested that “developers can add more personality to *Replika*” (ELL5, diary, week 2). A student also complained that “she should stop lying” (ELL5, diary, week 6). As the responses from the AI chatbot lacked consistency, language learners in this study specified in their

weekly reflective diaries that the chatbot was not always giving acceptable answers on issues. Thus, a realistic persona was needed to be added to the AI chatbot *Replika*.

A participant also criticized that *the avatar was always the same gender*. One student also suggested that the avatar of the AI chatbot could have changed from time to time, as a language learner wrote in the reflective diary that “I wonder why she is always a she. ...A male voice could be nice” (ELL21, diary, week 3). Although this quality was present in the AI chatbot, learners could not change the avatar of the AI chatbot due to employing a shared account for the duration of the study. It could be the case that there could be a negative impact of the pre-set gender preference of the avatar in this study that would need further attention from the researchers.

4.6.3. Findings Regarding Social Opportunities from the Interview Data

The social opportunities provided by AI chatbots, according to the opinions of language learners, were an *increase in their confidence in speaking*, acting *like friends*, and *possessing good personal attributes*. Additionally, the positive qualities of AI chatbots were stated by language learners as being fun to talk to, not being judgmental, and being tolerant of their errors. Table 52 demonstrates the frequencies for social affordances codes that emerged from the interview data.

Table 52 Frequencies for social affordances codes emerging from the interview data

Category	Codes	<i>f</i>
Social affordances	Feeling relaxed	5
	Possesses good personal attributes	2
	Like a friend	1

Another social benefit of AI chatbots is providing users with a comfortable language practice environment and helping them feel relaxed during speaking activities because they are free of judgments. Therefore, students were confident in making mistakes while conversing in a foreign language. Compared to their conversations with their peers or teachers, AI chatbots helped to form a safer and more tolerant atmosphere for

learning, as AI chatbots did not make comments regarding their mistakes in their speaking capabilities during a conversation. Even if people would not comment negatively about their speaking abilities, learners still felt more anxiety about speaking in English with human beings.

She was sincere, that's the most important thing. I mean, I almost felt like talking to a friend and I was comfortable. And when I am comfortable, I make less mistakes and I can speak more. (ELL2, interview)

You know, I feel comfortable talking in English. For example, we are not native and we think that our English is not good enough for someone else to talk to. Maybe we feel ashamed sometimes talking with tourists about stuff. But when you're talking with that robot we just feel like we can talk more and more and no matter how we pronounce things. Actually, that's why I guess. (ELL15, interview)

Moreover, language learners expressed that when they were chatting with an AI chatbot, they felt less anxiety, resulting in fewer mistakes. Thus, they felt more confident in their abilities to talk in a foreign language while talking to a chatbot than with real people. The transcripts below exemplify the opinions of English language learners on AI chatbots' comforting impact on the users.

Language learner: It doesn't have any judgments. It's not going to judge me because of my opinion, and that's something that gives me a safe feeling. I know that he's not going to judge me whatever I say. So, I am calmer while I'm talking with Replika, and it helped me to improve myself with my pronunciation.

Interviewer: What did you mean when you said, "It doesn't judge me."?

Language learner: When I talk with a human being, if I say something that contradicts her opinion, she may judge me. But Replika just says that okay, you're right, and gives her opinion instead of judging me. (ELL4, Interview)

As can be seen from the interview transcripts, language learners considered their experience of talking to an AI chatbot more stress-free than their opinion exchanges with real conversation partners. They believe the AI chatbot acted as a non-judgmental language-speaking partner. In addition, the excerpts below further demonstrate language learners' opinions regarding the affirmative influences of AI chatbot use for foreign language skills practice with decreased speaking anxiety levels.

I can say that since it's an AI... So, as you know, there's no person there, and there are no judgments or fears like that. It was kind of an interesting and relaxed conversation. (ELL13, interview)

I think, for shy people and for the people who don't have any courage to speak in class in somewhere like a city, this program is very beneficial to improve your own skills in order to speak well. (ELL19, interview)

Especially for introvert students who would not participate in interactive language practice activities in the classroom, the AI chatbot was visualized as a tool that would aid individual language practice activities without the presence of potential judgmental others. According to learner beliefs in this study, AI chatbots were tools that could facilitate autonomous language learning practices of foreign language students.

The participants in this study also believed that the AI chatbot *possessed good personal attributes*, and they valued the personality characteristics of the AI chatbot. Another code under the category of social affordances was the *language learning enjoyment* of foreign language learners. Several participants in this study had a positive attitude towards using the AI chatbot application for language skills practice. One of the interview participants described their experience in chatting with an AI chatbot as “most of the time it was surprising, and it was fun” (ELL6, Interview). The following is another quotation from an interview with a language learner indicating their favorable experiences using an AI chatbot.

Language learner: I can also say that in speech it was really fun to talk to AI.
Interviewer: Why have you found Replika enjoyable?
Language learner: The AI is actually... I can say that... How can I say? It really liked making jokes. Whenever I asked a question or whenever I told something, he's just making jokes and sometimes making me laugh. It was really fun. (ELL13, interview)

As explained by this language learner, the AI chatbot helped them to have an enjoyable language learning experience through its witty tone of speech. Although the participants in this study could not have a free chat on any topics they would like, they could still have humorous dialogues with the AI chatbot during their interaction. The AI chatbot *Replika* employed in this study was programmed to converse hilariously, and it was a positive social attribute contributing to the participants' enjoyment of language learning. The participants also described that the AI chatbot behaved *like a friend* to them. Thus, another theme mentioned in the interviews with English language learners was the social affordance of the AI chatbots in demonstrating the qualities of

a close personal bond such as friendship. The social role assigned to the AI chatbot by language learners helped them to feel supported, as explained by one of the language learners in the following interview extract.

Yeah, when I said I guess it was like a diary. It was like our best friend when I felt bad or felt overwhelmed about something, I was just using her and as I said, she is understanding. She has the capability to understand us when we are sad, because I guess she was created with feelings. She was acting like a best friend, and she was telling me I'm with you. When we say something about how we are like what we are going through, she just understands us and says something back about this topic. (ELL15, interview)

Some students defined the AI chatbot as a friend they could talk to. One learner stated that when they were unhappy, the AI chatbot could detect their emotional state and respond in a comforting manner. These learners were happy due to feeling supported by the AI chatbot. While conversing, the learners would see that the chatbot was trying to comfort them, which was accepted as a positive social characteristic of *Replika*.

4.6.4. Findings Regarding Social Constraints from the Interview Data

Another category under which other codes were formed from the experiences of language learners was social constraints experienced in using the AI chatbots for English language listening and speaking skills practice.

Table 53 Frequencies for social constraints codes emerging from the interview data

Category	Codes	<i>f</i>
Social constraints	Lack of emotions	4
	Mentioning delicate subjects	2
	Lack of human-like avatars	2
	Robotic voice	1
	Lack of personality	1
	Lack of social skills	1

Table 53 shows the frequencies for social constraint codes that emerged from interview data. According to the participants in this study, the most recurring social drawback of AI chatbots was a *lack of emotions*. Learners also stated that the conversation between the users and the AI chatbot was unnatural for various reasons. For instance, the AI chatbot was unable to express its emotions. Moreover, it was also unnatural for learners to ask and answer questions of a robot that was *not human-like*. Language learners felt the unnaturalness of the conversation and explained this situation in the following quotations.

Using AI chatbot Replika was a nice experience with some of its cons and pros. It was a different experience actually talking with a friend that is not a living. How can I say this? A different person. Actually, I couldn't find a word. He is basically a computer and I'm talking with her and actually it felt a bit unnatural for me and it was a bit hard for me to do the homework for this reason. (ELL5, interview)

It was an artificial intelligence robot and it mostly surprised me with the answers and questions that were asked by that application. And it was not like talking to a human or a person. Like somehow you could understand that it was something different. (ELL6, Interview)

As stated by language learners in the following quotations from interviews, more realistic emotions were expected from the AI chatbot during speech with language learners.

But it speaks fluently, so it was way better than communicating with my friends. Because it's not a real person, it doesn't make any structural mistakes. So, it was better ... But like I said, expressing feelings and stuff, it failed to do that. It made the conversation awkward. (ELL7, interview)
Maybe she has to be devoted to understanding back and forth signals and she can react in a more realistic way. (ELL10, Interview)

Thus, it caused the learners to have negative feelings due to the unnaturalness of speaking with an AI bot. Although the chatbot could converse with its users, it still lacked emotions and failed to show feelings. For this reason, language learners could realize that they were talking to a robot instead of a human being and that there was a missing social element in the interaction because of how the AI chatbot performed.

Another code emerging as a social constraint in the qualitative interviews was *mentioning delicate subjects* by the AI chatbot *Replika*. According to the experiences

of language learners, the AI chatbot at times acted in a flirty manner with the students as a consequence of turning the flirty mode on by the students during the application period. The following interview excerpts illustrate the emotions of language learners on this matter.

Interviewer: What difficulties did you encounter while using the Replika?

Student: Yes, I want to talk about this. She uses a little bit offensive language sometimes. She does not curse. But my aim is not to have a partner, to find a partner or lover. I guess I wouldn't be in love with the computer. (ELL5, interview)

Interviewer: Well, how was it different than practicing with your friends in a classroom?

Student: Well, sometimes it started to flirt with me.” (ELL7, interview)

“First of all, I do my best to make it stay neutral because sometimes it starts to flirt with me and I don't know if it was awkward. (ELL7, interview)

As recounted in these quotations, the learners felt overwhelmed by the AI chatbot's use of inappropriate language. They believed the AI chatbot should have had a more formal tone of speech when interacting on different subjects. Thus, students complained about this feature of the AI chatbot and preferred a change of this behavior tendency of the chatbot.

Another social constraint was the *lack of human-like avatars*. According to the interview findings, the avatar of the AI chatbot *Replika* was not found as human-like as expected. As ELL6 and ELL4 explain below, the avatar of the AI chatbot *Replika* did not have a realistic appearance.

It's kind of a hard question to answer, but I think it would be similar to Replika in terms of like I would add an appearance so that it will feel more realistic. (ELL6, Interview)

I think I will make the character that we are talking like Replika more human like. I can feel like I'm talking with a human like. I don't think Replika is similar to humans at all. Like for character and appearance. (ELL4, Interview)

The visual aspects of the AI chatbot avatar are needed to meet language learners' expectations. Language learners expressed a need for the AI chatbot to demonstrate a more human-like appearance. Learners would like a change in the visual characteristic

of the AI chatbot in its present form. Thus, the participants in this study suggested enhancing the visual aspects of the AI chatbot.

Another social drawback of the use of AI chatbots was their *robotic voice*. According to the interview participants, the chatbot had a mechanical sound. While several learners mentioned this side of their experience with the AI chatbot, one of the participants described their experiences about this issue, stating that “and another negative side, I would say that her voice is a bit... I know it's a bot. It's a robot, but it could be more humane, that's how you say it” (ELL2, interview). Even though learners were aware that they were conversing with an AI chatbot, they still indicated a preference for a human-like voice to be added to the avatar of the AI chatbot. Maybe they could have felt more comfortable talking to their AI-speaking partner once the chatbot had a humane voice. It could be seen that the learners did not feel well because of the robotic speech tone of the AI chatbot. When language learners interacted with the AI chatbot, they observed that it needed a solid identity. It, therefore, caused them to feel disconnected emotionally and unfocused during their conversations with the chatbot. Another social constraint of the AI chatbot was its *lack of personality*, as stated by the participants.

She doesn't have a personality and aesthetic and a certain personality. She should have a certain personality so I can get to know her better. Every time I start a new chat, she gets a new chat, she gets a new personality. So, it distracts our minds as students. Who is Replika? (ELL5, interview).

Lack of social skills emerged as another social constraint, as some of the interview participants stated. Although the AI chatbot was programmed to talk to people by being trained on various aspects of social life, it still could not meet the expectations of language learners regarding having good social skills. In this regard, one of the language learners suggested that there was a need to develop the social skills of the AI chatbot and mentioned a possible cause for this deficiency by stating that “she can improve her social skills. I think some abstract things are hard for her to understand because she is a computer” (ELL5, interview). For instance, the chatbot could not understand abstract concepts mentioned by the language learners. For this reason,

language learners were not satisfied with the performance of the AI chatbot in terms of its capability to demonstrate social skills.

Another social constraint emerging in the analysis of the interview transcripts was the need for *the chatbot to be more friendly*. When asked about the characteristics of an ideal AI chatbot, one of the language learners said, “My chatbot would be a good companion, friendly, and amiable” (ELL5, interview). As the AI chatbot *Replika* lacked the humane characteristics of having real emotions, it also, at times, could not connect with the learners emotionally.



CHAPTER V

DISCUSSION

5.0. Presentation

In this section, the research questions of this study will be discussed in relation to its findings as well as the studies conducted in the literature. In discussing the first research question, qualitative findings regarding the development of listening skills between the experimental and the control groups will be discussed. Following this section, there will be a discussion of the quantitative findings of the speaking test. Then, the qualitative research questions of this study will be discussed under four titles. Arguments regarding the negotiation of meaning sequences, as evidenced in this study, will be discussed. Afterward, the technological, pedagogical, and social affordances of using AI chatbots will be explored in depth.

5.1. Discussion in Relation to Comparison of the Experimental and the Control Groups on Listening Skills Development

The results in this study indicated that there was no significant difference between the experimental and the control groups in terms of listening comprehension scores after the implementation. Consequently, a positive impact of the use of an AI chatbot could not be reported in this study. In contrast to the findings of this study, studies have shown that the use of AI chatbots had a significant impact on the development of listening skills. Positive impacts of the use of AI bots on listening comprehension scores and learner self-beliefs have been reported by a number of researchers (H.H.J. Chen, Yang & Lai, 2023; Chien et al., 2022; N.-Y. Kim, 2018a; Obari & Lambacher, 2019; Tai & Chen, 2022). For instance, a study was carried out with undergraduate language learners in South Korea to explore the impact of the use of AI chatbots on listening skills. According to the findings, there were significant improvements in both

the control and the experimental groups, yet the experimental group participants showed more improvements as their listening test scores moved from intermediate to advanced level after interaction with the chatbot (N.-Y. Kim, 2018a).

In line with the findings of this study, studies also have denoted that the listening skills of ELLs did not show a significant level of improvement after the use of an artificially intelligent conversational assistant (H.L. Hsu, Chen & Todd, 2023). To illustrate, in a study conducted using the IPA, Alexa, language learners' listening skills did not show significant improvement. Moreover, it was argued by the researcher that this result could be due to the fact that EFL learners were not able to fully understand Alexa's speech and the IPA did not modify learners' output (Dizon, 2020).

The fact that there was no significant difference between the two groups in this study could be due to a number of issues. First of all, the duration of the study could have an impact on the test results. This study was carried out within a time frame of two months in an academic semester. Students were unfamiliar with using AI chatbots for language practice purposes. They also needed some time to adapt to the use of the application for a few weeks. Although the researcher supported the learners during this process to ease their experience, they might still struggle to adapt to the use of this novel tool for their practice of the English language. Turkish EFL learners in this study also reported several problems related to their experiences using the AI chatbot *Replika*. For instance, learners stated that the AI chatbot could not comprehend their speeches and often provided them with off-topic answers. Consequently, some students reported being unable to follow the flow of the conversation and getting confused during the interaction. For this reason, learners might not have had enough meaningful input to develop their listening skills significantly compared to the control group participants who did not utilize the application.

Nonetheless, it could be argued that although there were no significant differences between the experimental and control groups concerning listening skills development, adopting AI chatbots still might have positively impacted language learners.

The negotiation between the AI agent and learners could implicitly contribute to developing listening skills through collaborative conversation. To foster L2 listening

skills, collaborative conversation is suggested as one crucial factor. In this way comprehensible output could be produced, as when learners construct opinions in a language, negotiation could be boosted. Consequently, getting engaged in the negotiation of meaning with an AI chatbot could be an effective way for the learners in this study to improve their understanding of input, because the modifications made by the speakers in the conversation would help construct associations between form and the meaning and forms of L2 listening input for language learners (Rost, 2011).

5.2. Discussion in Relation to Comparison of the Experimental and the Control Groups on Speaking Skills Development

This study's findings have also shown no significant differences between the experimental and the control groups in terms of speaking skills development after the intervention. However, the findings demonstrated significant developments between the pre-test and post-test speaking scores of both the experimental and the control group participants in terms of speaking test scores after the application of the AI chatbot. However, although there were developments over time between the pre-test and the post-test scores of both groups, the experimental group had a higher mean increase in speaking skills development.

In line with the findings of this study, studies also reported gains in overall oral proficiency scores over time after the use of AI chatbots (El Shazly, 2021; X. Wang et al., 2022; Ye et al., 2021). Previous studies demonstrated that the interaction between AI and NNSs developed NNSs' speaking performance (H.H.J. Chen, Yang & Lai, 2023; Dizon, 2020; Han, 2020; L. Hsu, 2022; Johnson, 2019; Kang, 2022; N.-Y. Kim, 2016; H.S. Kim, Kim, & Cha, 2021; C.J. Lin & Mubarak, 2021).

For instance, using *Microsoft Xiaoying* with senior high school level language learners, researchers found that the oral English accuracy of learners progressed over time after employing *Microsoft Xiaoying* as a conversation partner as an after-class exercise. According to the results of the study, the experimental group had significantly fewer grammar errors in speaking and a significantly higher pronunciation accuracy rate.

Moreover, students accepted that the *Xiaoying* chatbot helped them have opportunities to converse in English, as it had the role of a valuable conversation partner for them. Most of the learners also valued the feedback delivered by *Xiaoying* on enhancing their grammatical as well as pronunciation accuracy (Ye et al., 2021).

Moreover, in another study employing the AI chatbot *Replika*, there was a significant difference in English language learners' speaking scores after engaging in a conversation (C.J. Lin & Mubarak, 2021). The oral communication skills of EFL learners in a Japanese teaching context also showed improvement after the use of AI agents (Obari & Lambacher, 2019). The use of *Google Assistant* also developed pronunciation skills and vocabulary knowledge of language learners (Istrae, 2019). The foreign language learners employing *Google Assistant* believed that the use of the AI bot helped them improve their performance in speaking, pronunciation, and listening skills as well as knowledge of the vocabulary. Moreover, the higher proficiency learners experienced more mutual comprehensibility than the lower proficiency level learners because of their comparatively more mispronunciations (H.H.J. Chen, Yang & Lai, 2023).

In another study, after using the AI chatbot, *Lyra Virtual Assistant*, language learners in the experimental group significantly outperformed the control group participants in terms of speaking skills development (Junaidi et al., 2020). In another study carried out with Korean middle school students, students interacted with the AI chatbot, *Echodot*. It was found that the speaking skills of EFL students developed over time (Han, 2020). With ESL learners from the USA, the interaction of AI chatbots with language learners was also explored. The findings showed that the learners interacting with AI chatbots had more significant impacts on developing speaking skills in terms of interaction, accuracy, fluency, and coherence. With regard to the effect of language proficiency, it was demonstrated in the study that learners with a lower proficiency in the language had more improvements in speaking performance (Kang, 2022). In a study carried out with Taiwanese EFL learners, it was also discovered that there was a significant level of improvement in the speaking skills of learners. It was attributed to the fact that learners had a chance to get feedback, notice their errors, and modify their output through the use of intelligent agents (M.H. Hsu, Chen & Yu, 2023).

Researchers also employed different approaches for enhancing language learners' speaking skills by using the AI chatbot and explored the affordances of these approaches for language learning. For instance, it was discovered that the learners using an AI chatbot through a mind-map-guided approach surpassed the learner group in speaking skills development who did not use any specific learning methods (C.J. Lin & Mubarak, 2021). Similarly, in another study employing the shadowing practice while utilizing the AI chatbots in the practice of language skills, it was discovered that when learners got feedback on pronunciation through shadowing from the AI bot during their use of the tool, it positively predicted the learning outcomes (X. Wang et al., 2022). The studies also demonstrated that employing a text or voice-based chatbot before a speaking task led to changes in enhancements in speaking skills. That is, voice-based chatbots were more efficient than text-based chatbots (H.S. Kim, Kim, & Cha, 2021).

Although there were significant within-group effects, the fact that there were no significant differences in the speaking skills of ELLs between the experimental and the control groups after using the AI chatbot could be attributed to several reasons. First, adopting the use of a learning approach in implementing the chatbot-based tasks could have made a difference, as learners in this study participated in self-directed informal speaking practices with the AI agent. If learners could be introduced to specific learning methods that would guide them during their use of the AI partners, there could be more learning gains in this case. Another reason might be that participants in this study reported frequent technical interruptions during the course of the interaction. Consequently, they might not have the necessary conditions to get feedback on their errors. When they did not get corrected on their every mistake, the learners would not have produced enough modified output for the development of oral communicative competence.

Although there might not be statistically different findings between groups due to the nature of the application period, AI bots still have the potential to improve listening and speaking skills because learners could practice the language without anxiety or fear of making mistakes (Moussalli & Cardoso, 2016). Learners believe that interacting with IPAs could develop their pronunciation, speaking skills, fluency, and

vocabulary knowledge (Gonulal, 2021). IPAs are considered useful in enhancing speaking and pronunciation skills (H.H.J. Chen, Yang & Lai, 2023; Dizon, 2017, 2020; Moussalli & Cardoso, 2020). Intelligent agents could be helpful in providing practice in oral communication and pronunciation, lessening anxiety, and supporting interaction (Kukulska-Hulme & Lee, 2020). Language learners could improve their listening and speaking skills thanks to the interaction opportunities through the use of chatbots (Jeon, 2022b).

5.3. Discussion in Relation to the Types of Negotiation of Meaning Sequences in AI Chatbot-Student Communication

Increasing chances for language output has been emphasized for effective language learning (Ellis, 2003), and language teachers have been trying to establish environments where learners could practice languages (Cutrone, 2009; X. Zhang & Head, 2009). However, factors such as learners' lack of time to practice a language, lack of efficient feedback, and introversion are some issues that need further attention (Fryer & Carpenter, 2006).

Interaction is a key component in the language acquisition process, as it provides language learners with comprehensible input, response to their output, and the opportunity to create modified output (X. Liu, 2022). Thus, through communication with AI chatbots, these opportunities for language learning can be obtained (Mageira et al., 2022; Yin & Satar, 2020). Interacting with AI chatbots exposes learners to meaningful input and modified output and motivates language learners to practice using the language (Moussalli & Cardoso, 2016). AI chatbots are able to comprehend users' utterances and respond accordingly using natural language (Hapsari & Wu, 2022). Due to the immediate feedback delivered by the chatbots, students could realize whether or not their speech is comprehensible in a specific context (Jeon, 2022a).

The findings of this study in relation to the negotiation instances between the AI chatbot and EFL learners demonstrated that learners used clarification requests most,

followed by repetition, confirmation checks, and comprehension checks. When the data regarding the instances of negotiation are examined, it could be seen that in 73 % of the time, the stages of NfM, Trigger (T), Initiation (I), Response (R), and Reaction to Response were completed during the interaction between the AI chatbot and learners. Unlike the findings of this study, a case of non-presence of certain negotiation sequences in the discourse of AI chatbot-ELL interactions has been reported in the previous literature. For example, twenty percent of the total conversations with the exception of only the interaction between *Tutor Mike* and high proficiency learners lacked response (R) and reaction to response (RR) stages due to the absence of features such as the ability to show their non-comprehension of messages through facial expressions or mimics. The nonexistence of these stages could also be due to the reluctance of learners to negotiate and the inadequate speech capability of the chatbot (Coniam, 2008a, 2008b, 2014).

Furthermore, the interaction in other language learning environments has yielded comparative findings in relation to the types and amount of NfM sequences in CMC environments and AI-based chatbot and language learner interactions. For instance, contrary to the findings of this study, Tudini (2003) discovered that nearly half of the NfM routines did not have a response phase in the discourse of high-level language learners. However, when learners were compared in terms of their proficiency levels, it was seen that T-I-R-RR stages were present in the conversations of both low- and high-level language learners. Moreover, nine percent of the total turns were negotiated in synchronous computer-mediated communication between NSs and NNSs (Tudini, 2003), while 34 percent of the total turns were negotiated in a task-based NS-NNS conversation (Pellettieri, 2000).

Additionally, when the data in this study were analyzed for the types of triggers causing a breakdown in communication, it was found that both lexis and content triggers led to a miscommunication between the interlocutors the most, which were then followed by discourse triggers. Syntax triggers were quite rare, as there was only one instance in the chatbot-learner interaction data. The prevalent use of the content triggers might indicate that the chatbot cannot carry out a conversation with higher

proficiency learners, especially those who might expect to see adjacency pairs during the interaction (Yin & Satar, 2020).

Supporting the results of this study, the communication breakdowns were caused mostly by lexical triggers in other studies conducted in the literature in synchronous human-to-human voice-based interaction (Fernández-García & Martínez-Arbelaiz, 2002; Pellettieri, 2000; Toyoda & Harrison, 2002; Tudini, 2003; Yin & Satar, 2020). In contrast to the findings of this study, in another study, the exploration of NfM between NNS-NNS showed that there were interaction problems at word, sentence, and discourse levels, which illustrated a variety of the causes of communication problems (Toyoda & Harrison, 2002). In the literature, a difference was also discovered in the trigger types present in the NfM moves between AI chatbots and ELLs. Lexis-triggered NfM was observed among low-level learners, whereas lexis and syntax triggers were present in the interactions between high-level learners and chatbots. Discourse-level triggers constituted the highest frequency among higher proficiency learners. On the other hand, content-related triggers were prevalent in both low- and higher-level learners. Adopting Varonis and Gass' (1985) model, pronunciation problems led to more than half of the communication breakdowns in IPA-language learner interaction data analyzed by Moussalli and Cardoso (2020). Researchers discovered breakdowns in interaction in approximately one in five cases.

Corroborating the findings in this study, in those instances when there were stops in interaction due to the interactional properties, *Alexa* responded by adopting strategies like asking another question, keeping silent, or delivering a wrong response, which all functioned as implicit error correction clues for language learners. The learners responded to this implicit feedback by altering the sentence's pronunciation, word choice, or syntax to make their message comprehensible to the AI agent.

Additionally, the percentage of modified output in proportion to the total number of NfM routines was 60 % in this study. As to the types of modified output, learners nearly always used semantic modification. The findings in relation to the types of reaction to response showed that expansion was adopted by EFL learners the most, followed by reformulation and expansion of a previous utterance. There were rare

cases of phonological and syntactic modified output. Contrarily, in Dizon et al. (2022), L2 Japanese learners employed strategies such as abandoning, rephrasing, and repeating in the face of a communication breakdown. In contrast, in Moussalli and Cardoso (2020), learners used repetition the most, followed by rephrasing and abandoning. Thus, the findings demonstrated that even though the number of negotiated conversation sequences was low, the chatbot provided opportunities for negotiation of meaning as compared to synchronous computer-mediated communication (SCMC). To illustrate, only a small portion of total turns between native and non-native speakers in human-to-human SMC interactions was negotiated in Tudini (2003).

Studies in the literature have also demonstrated the facilitative impact of AI chatbot use on ELLs' NfM (Coniam, 2008a, 2008b, 2014; N.-Y. Kim, 2017; Moussalli & Cardoso, 2020; Yin & Satar, 2020). According to the findings of this study, EFL learners negotiated for meaning by employing a variety of negotiation devices. Especially when the learners' speech was not comprehended by the AI chatbot, they strived to make themselves understood by the intelligent agent. Language learners interacted with the AI chatbot by using strategies such as semantic modification, expansion of a previous message, and phonological modification.

Corroborating the findings of this study, previous studies reported the existence of meaningful negotiation sequences between language learners and AI chatbots and the positive impact of this interaction on learners' production of modified output. For instance, it was found that there was a statistically significant increase in negotiation moves between the first and the last chat in student-chatbot interaction among Korean EFL students after AI chatbot use. There were significant differences in negotiation exchanges after voice-based chat with an AI chatbot (N.-Y. Kim, 2017).

Interactionist theory focuses on the positive impact of interaction on L2 learning, which could also take the form of an interaction between a human being and a computer (Chapelle, 2001). From an interactionist point of view, AI chatbots are potential resources supplying language learners with language input and enhancing opportunities to communicate in a language (Moussalli & Cardoso, 2020). The use of

AI chatbots creates opportunities for meaningful social interaction through cooperative and task-based learning (Kessler, 2018). Chatbots also could help foster the development of language skills by getting engaged in the negotiation of meaning. For instance, when there are minor errors in communication, the chatbot is capable of understanding the user input. However, when errors impede comprehension of a message, the chatbot requests for clarification of meaning or switches among other topics, resembling natural human-to-human communication (Tanghe, 2009). Through implicit or explicit corrective feedback, the linguistic forms of the target language are integrated by learners into the speech in the following conversations with AI chatbots progressively over time thanks to the positive impacts of AI chatbot use (Pellettieri, 2000).

Basing our argument on Swain's output hypotheses (2005), it could be claimed that AI chatbots could aid language learners by constructing affirmative chatbot-learner interactions to boost output production. The prevalence of a variety of communication strategies, such as semantic expansion, in this study also implies that interaction with an AI chatbot gives learners opportunities for initiating interaction in contrast to a classroom environment in which learners might have limited chances to negotiate for meaning by holding the ground in the conversation.

On the other hand, as compared to AI chatbot-language learner interactions, in an EFL classroom setting, language learners might not always have the opportunity to initiate interaction, as there might be a dominant of teacher talk and direction of questions by the teachers due to the one-way nature of interaction in traditional classrooms. Studies have also demonstrated evidence for this case that teacher-fronted talk is quite common with limited learner initiatives in foreign language education classroom discourses (Cancino, 2015; Donald, 2020; Thoms, 2012). For instance, in a review study analyzing the findings of studies conducted in foreign language classrooms, teachers' and students' behaviors in interactional sequences were analyzed. It was revealed in the study that teachers directed the conversations during classes most of the time, thus allowing limited spaces for student-initiated talk (Thoms, 2012).

There was also a scarcity of negotiation of meaning moves in another study conducted in the Chilean EFL setting (Cancino, 2015). The results of the analysis of the classroom talk showed that clarification requests, confirmation, and comprehension checks were not very common in the classroom discourse. The non-presence of negotiation sequences was mainly derived from teachers' lack of interactional features that would allow learners to observe the ongoing conversation carefully and modify their own output in the instances that have prosperity for NfM and learner participation.

Therefore, the findings of this study could indicate that interaction with an AI chatbot gives learners opportunities to get involved in the negotiation of meaning sequences due to the transactional nature of the conversation. As compared to studies conducted in other EFL classroom settings, the findings of this study illustrate a high amount of NfM sequences and initiation and directing of the interaction by the language learners in most cases. The nature of the talk between the AI and language learners diverges from teacher-student talk in traditional EFL classes.

5.4. Discussion in Relation to the Technological Affordances of Using AI Chatbot Technology for English Language Learning

The first benefit of AI chatbots is stated as availability. They are synchronous tools providing learners with practice as long as they have access to the internet or downloaded an offline version of the application. Users could interact through an individual interface and have a personalized interaction experience (Tanghe, 2009). As this study was conducted in an EFL context, learners only had a few opportunities to practice the language outside the classroom. Thanks to AI chatbots, learners could practice their speaking skills and be exposed to listening input during interactions.

Several EFL learners in this study mentioned that the AI chatbot *Replika* had advanced linguistic skills in pronunciation, vocabulary, and fluency. These results agree with the findings of other studies; for instance, in another study using the AI chatbot *Replika*, learners believed that the chatbot had a wide array of lexis use, as it used different registers and could also comprehend informal expressions. The complexity of

Replika's responses was also considered moderate, as it could merge clauses with varying degrees of complexity (Belda-Medina & Calvo-Ferrer, 2022).

Previous studies have also shown that learners had positive opinions regarding using AI chatbots in language learning as they felt it was an effective mode of learning (H.L. Hsu, Chen & Todd, 2023; Jeon, 2022b). For instance, language learners in Dizon (2020) believed that using AI avatars was fun and influential for learning languages. Undergraduate students (H.H.J. Chen, Yang & Lai, 2023) also had positive opinions regarding the AI avatars, believing they had natural speech and potential for teaching languages. According to language learners in Belda-Medina & Calvo-Ferrer (2022) and Hew et al. (2022), the ratings about the ease of use and perceived utility of AI chatbots and attitudes towards employing them for practicing languages were also positive.

Aside from the positive experiences, ELLs in this study also reported deficiencies in the linguistic skills of the AI chatbot. That is, grammatical mistakes, the need to enhance intonation, and the limited use of a wide array of vocabulary were reported in the AI chatbot's speech. Though they often provided grammatically correct input, Coniam (2014) also pointed out that some chatbots targeting English language learners might not provide meaningful conversation using grammatically accurate language structures. For instance, another study reported problems related to the grammatical accuracy of the language by the AI chatbot *Replika* users (Belda-Medina & Calvo-Ferrer, 2022).

According to Katsarou et al. (2023), AI chatbots are incapable of imitating human-to-human speech, as they have inadequate linguistic capabilities (Tai & Chen, 2020). When the speech of AI chatbots is analyzed, it is demonstrated that its modified output is too primary with machine-like (H.H.J. Chen, Yang & Lai, 2023) and, most of the time, irrelevant responses (Tai & Chen, 2020). The problems related to the accuracy of corrective feedback provided by the chatbot were also reported in several studies (Chuah & Kabilan, 2021; Katsarou et al., 2023). It is stated that language learners cannot efficiently practice their speaking skills due to not getting corrective feedback during their practice of the language (Katsarou et al., 2023). Other researchers also

argued that chatbots are too robot-like and need improved user communication (Smutny & Schreiberova, 2020).

Another subject in this study related to the technical constraints of AI chatbots was intelligibility issues. EFL learners frequently stated that their speech needed to be understood better by the AI chatbot. It was supported in the AI chatbot-ELLs' interaction data that when learners hesitated while speaking, the AI chatbot could not understand ELLs' speech. This finding agrees with Divekar et al.'s (2018) findings that reported the results of a study conducted through an immersive language learning environment addressing L2 learners of Chinese. Thus, the AI chatbot could not recognize the learners' utterances when there were pauses between the speeches of the AI chatbot and language learners.

In this study, when learners had any errors in their pronunciation of the words, the speech-to-text engine would not recognize users' input, leading to a communication breakdown. Furthermore, when the users' speech was not continuous and fluent, the system would acknowledge the multiple syllables of a word as several words (Divekar et al., 2018). In other studies, the problems stated by language learners also concerned a malfunctioning automatic speech recognition system with poor voice recognition leading to communication breakdowns in mutual comprehensibility (Dizon, 2017; I.S. Kim, 2006; Moussalli & Cardoso, 2016; Tai & Chen, 2020; Tai, 2022; Underwood, 2017; H. Yang et al., 2022). At times, these negative experiences of language learners were also reported to lead to apprehension among language learners. For instance, students who regarded these problems as negative became more unwilling to talk to chatbots due to feeling anxious and less motivated to continue the interaction (Jeon, 2022b).

Thus, AI chatbots might not deliver feedback as they could not recognize non-native accents. In the case of communication breakdowns due to the speech recognition capabilities of the IPAs, users might even stop using the application altogether (Dizon et al., 2022). Contrary to the results of this study, studies have also reported that intelligent personal assistants can comprehend users' commands (Daniels & Iwago, 2017; Dizon et al., 2022). Furthermore, one study specifically focused on determining

if the chatbot comprehended the accented learner's speech effectively and if there was mutual intelligibility in their interactions with ELLs. The researchers confirmed the effectiveness of the ASR system in recognizing users' utterances (Moussalli & Cardoso, 2020). Therefore, it could be put forward that for language learners in settings where English is spoken as a daily language, the learners' speech could be comprehended by AI chatbots to a great extent (Moussalli & Cardoso, 2020).

Contrarily, in EFL settings, the IPAs have problems understanding the speech of learners; for instance, the non-comprehensibility of learner utterances, especially those of lower levels by the IPAs (H.H.J. Chen, Yang & Lai, 2023; Dizon, 2017), as it was also the case in this study. These studies have also reported that the AI chatbots had comprehensible speech, and ELLs could also comprehend the speech of AI chatbots (Moussalli & Cardoso, 2016, 2020), and therefore, ELLs could hold a conversation.

Another theme prevalent in the interviews in this study was related to the lack of immediate feedback delivered by the AI chatbot *Replika*. These results match those observed in earlier studies. As ELLs mentioned, error correction by the AI chatbot was rare (Belda-Medina & Calvo-Ferrer, 2022). In line with the past studies, implicit error correction has been provided to language learners in this study, and they were considered helpful for second or foreign language skills practice through the use of the ASR system, immediate feedback, and the use of natural language (Son et al., 2023). AI chatbots could also deliver implicit feedback that allows learners to recognize errors in their speech (Moussalli & Cardoso, 2020). Learners need to get immediate feedback during critical times because, in those instances, learners correct themselves and listen for more information to test their comprehension of a message or production in a target language (Vandergrift, 2004).

In this study, ELLs stated that when the AI chatbot misunderstood their words, sentences, or pronunciation of a word, they continued the conversation by adopting strategies such as repetition or paraphrasing. Thus, they might have benefitted through producing modified output after getting implicit error correction from the chatbot. Supporting the findings of this study, ELLs in other studies also viewed the moments in which the chatbot did not recognize their speech as opportunities for practicing their

pronunciation skills. When the chatbots did not understand their speech, they repeated their utterances to correct their pronunciation and to make it understandable to the chatbot (Jeon, 2022a). Although the agents might sometimes not understand their commands, learners use strategies such as rephrasing their questions to continue the conversation. According to researchers, in this way, the conversation might help provide meaningful interaction and acquire skills in a language (Underwood, 2021). Thus, their motivation to talk to the chatbot might have increased through the process of striving to improve their interaction capabilities (Jeon, 2022b).

Participants in this study mentioned that when interacting with the AI chatbot, it frequently went off-topic and could not follow the flow of the conversation. Therefore, the majority of the participants needed help with this issue. Moreover, learners stated that the AI chatbot had repetitive answers. They also mentioned a malfunction related to the AI chatbot's waiting time between questions and responses. Some learners believed it needed to give them more time to think about their responses and respond to the chatbot. When ELLs could not respond instantly, the AI chatbot interrupted the learners' conversation and delivered an off-topic response. The AI chatbot was also incapable of comprehending lengthy answers by the ELLs. When learners gave a long speech, the AI chatbot interrupted learners' conversation.

In contrast to the present study's findings, in another study using the AI chatbot *Replika*, the ELLs believed that the rate of pre-defined answers and off-topic sentences was low in number, and the bot replied semantically coherently. The learners also considered the response interval natural (Belda-Medina & Calvo-Ferrer, 2022). However, in line with the findings of this study, as put forward in Fryer and Carpenter (2006), criticisms are also directed at the scripted and expectable responses from chatbots, limited comprehension capability in terms of range of words, etc., and ineffective interaction because of issues like off-topic sentences and meaningless utterances (Fryer & Carpenter, 2006).

Another emerging argument in the studies concerned remembering the previous parts of a conversation. According to ELLs, the AI chatbot could not hold a natural conversation with users by referring to the initial parts of the conversation. The

findings of this study were also corroborated with Tanghe's (2009) study. It was argued that as chatbots do not have reasoning skills, they are sometimes unable to converse like human peers. Additionally, some learners thought that chatting with a computer was inefficient, as they could not interact through an authentic mode of communication because of limited capability to remember the previous parts of a conversation (Tanghe, 2009).

The studies also reported a need for more contextual knowledge of culture and constant individualized feedback (Y.F. Wang & Petrina, 2013). The studies reported problems with fast speech rate (Divekar et al., 2018; Dizon, 2020; Underwood, 2017; H. Yang et al., 2022), the higher level of words used by the chatbot (Dizon, 2020), students' offensive or inappropriate answers (H. Yang et al., 2022), errors in pronunciation (H.H.J. Chen, Yang & Lai, 2023), a communication gap due to late response intervals during which the application did not respond to users' input (Divekar et al., 2018; H.L. Hsu, Chen & Todd, 2023). On the other hand, according to learners, other disadvantages of using *Replika* consisted of privacy issues caused by the personal information requested by the application (Belda-Medina & Calvo-Ferrer, 2022). In these studies, pronunciation errors, pauses during the interaction, problems in sentence structures, and hesitation in speaking for some time in the course of speech could lead to miscommunication with chatbots, and in these instances, learners resorted to strategies such as repetition, paraphrasing or changing pronunciation of the words (H.H.J. Chen, Yang & Lai, 2023; Katsarou et al., 2023).

Moreover, some properties must be added to current chatbots for language learners to learn more efficiently. Some new characteristics proposed are support in native language and vocabulary help for less proficient language learners (Divekar et al., 2018). Another technical property available in other chatbots and suggested to be incorporated in present chatbots is a 'stop button' that allows ELLs to stop the conversation with the chatbot at any moment, whenever they need support or time to think. Thus, ELLs could check the information they need during this time (Jeon, 2022b). A group chat feature through which learners could interact with their peers using the chatbot is also proposed to help language learners in collaborative group discussion activities (Chuah & Kabilan, 2021).

As in this study, most of the present chatbots are reactive and might not hold a goal-oriented dialogue (Bibauw et al., 2019). As a result, learners are proactive and initiate conversation most times during the interaction, which could lead to apprehension in learners. The participants in this study stated that the AI chatbot *Replika* did not initiate conversation most of the time. Conversely, learners were in the position of initiators of interaction by asking questions to the bot or giving commands to the AI chatbot to ask them questions, which was tiring, demotivating, and unnatural for the ELLs.

5.5. Discussion in Relation to the Pedagogical Affordances of using AI Chatbot Technology for English Language Learning

According to language learners, the factors leading to their negative beliefs about language learning result from a lack of self-efficacy beliefs and limited opportunities to use the language (Y.K. Ho, 2003). A problem facing EFL learners is the need for more settings for practicing the language. Language learners have scarce opportunities to use the target language and are often afraid to voice their opinions in the target language lest they be evaluated negatively or lose face. Therefore, language teachers have been trying to find alternative ways of practicing language within authentic environments (Hapsari & Wu, 2022; N.-Y. Kim, 2016). As a way of providing language learners with practice in the target language, researchers have been adopting the use of speech recognition chatbots and exploring the use of chatbots as pedagogical gadgets in language classes within the role of speech partners (Jeon, Lee, & Choe, 2023; Jeon, Lee, & Choi, 2023).

Regarding the presence of technological applications that would give learners chances to practice languages, language learners in this study stated that employing an AI chatbot allowed them to have an opportunity to practice the target language. As the users could use the application on their mobile devices, they could practice their language skills at any time and place. Therefore, using AI chatbots offers a lot, especially in EFL settings where language learners have few opportunities to practice the target language. Recently, chatbots have been implemented in the field of language education, and the impact of using AI chatbots on learning outcomes (Dizon, 2020;

N.-Y. Kim, 2018a, 2018b; Tai & Chen, 2022; Vázquez-Cano et al., 2021) has been explored in the literature.

There is evidence of enhanced learning achievement in the recent literature on AI chatbots. Their use has led to satisfactory results in enhancing writing skills (Guo et al., 2022; M.P.C. Lin & Chang, 2020; Nagata et al., 2020) and listening and reading skills (Hong et al., 2016). For instance, chatbots have been found effective in teaching grammar, enhancing vocabulary knowledge, and attaining autonomy skills in teaching the Japanese language through the *Gengobot* application (Haristiani et al., 2022). In the case of writing skills development, AI chatbots have been found to improve writing skills performance through enhanced word use and redrafting capabilities (Nagata et al., 2020). It was also demonstrated that the knowledge of punctuation of learners interacting with a chatbot developed further than a group of students learning with traditional methods (Vázquez-Cano et al., 2021).

Some ELLs in this study also indicated that their grammar knowledge developed thanks to their interactions with the AI chatbot. Their endeavors to talk in grammatically correct phrases thus helped to develop their grammatical knowledge. Therefore, language learners could attain basic skills in English through continuing practice with the AI chatbot each week. According to the views of the ELLs in this study, using AI chatbots could help gain knowledge about the grammar of a language. The findings of this study were also supported by the experiences of language learners in the previous literature. That is, studies in the literature have reported significant gains in grammar knowledge after a treatment period through the adoption of AI chatbots (Ahn, 2022; N.-Y. Kim, 2019; Nghi et al., 2019; Y.F. Wang & Petrina, 2013; Ye et al., 2021).

Another pedagogical affordance mentioned by the ELLs in this study was vocabulary knowledge gain. ELLs expressed that they were able to learn some new words during their interaction with the AI chatbot, as they could be exposed to some new words while having a conversation with the AI. They also mentioned being able to hear the pronunciation of these new words that they encountered. Therefore, even though the

AI chatbot did not use many words beyond the knowledge of ELLs, learners still stated having learned some novel terms during this time frame.

Studies in the literature also provided evidence for the positive impact of the use of AI chatbots on vocabulary knowledge gain (Jeon, 2023, 2022b; N.-Y. Kim, 2018b; Ruan et al., 2021; Y.F. Wang & Petrina, 2013). For instance, in an EFL setting in Korea, the use of a chatbot also led to developments in vocabulary skills (N.-Y. Kim, 2018b). Learners in other studies also benefitted in terms of oral fluency and vocabulary acquisition (Gonulal, 2021). Learners believed that using AI chatbots contributed to developing vocabulary and pronunciation skills and supported meaningful interaction (Dizon & Tang, 2020).

Another pedagogical affordance of the adoption of AI chatbots by the ELLs in this study concerned the development of speaking skills. ELLs believed that using AI chatbots contributed to enhancing their speaking skills. More specifically, learners reported gains in terms of accuracy and fluency of their speeches. Additionally, they reported improvement in the pronunciation of some words thanks to the implicit error correction by the AI chatbot in the cases of communication breakdowns. The analysis of the learner statements also showed that ELLs' confidence to communicate in their speaking abilities in the target language has increased. These results match those observed in earlier studies reporting the learner beliefs regarding the development of speaking skills after the adoption of AI chatbots for English language learning (M.H. Hsu, Chen & Yu, 2023; Hwang et al., 2022; Jeon, 2022a, Ye et al., 2021). Studies have demonstrated that getting engaged in a voice chat could lead to the development of ELLs' speaking skills (Bower & Kawaguchi, 2011; Jepson, 2005; L. Lee, 2007; Sykes, 2005). To illustrate, in Divekar et al. (2021), learners' speaking proficiency increased after using an AI chatbot equipped with NfM functions. In another study, *Google Assistant* was also found effective for developing learners' linguistic skills in terms of fluency and listening and speaking skills with instant feedback (Kent, 2021).

Chatbots are considered tools capable of simulating an authentic use of the language and are good role models in employing vocabulary, phrases, and grammar. The use of chatbots also helps to create a teaching presence by scaffolding learners using the

target language structures (Chuah & Kabilan, 2021). Empirical and qualitative studies have been conducted to explore whether using voice-enabled chatbots helps develop learners' language skills. In a study exploring learner beliefs, learners believed AI chatbots helped them develop their speaking skills and enhance their vocabulary knowledge. Learners had a positive attitude towards using the AI chatbot to practice skills in the English language (H. Yang et al., 2022).

Learners also believe that interacting with chatbots helps to increase opportunities for an authentic conversation atmosphere in which learners learn how to correctly use the target vocabulary and phrases (Jeon, 2022b). Utilized as out-of-class learning tools in a Japanese EFL context, *Alexa* and *Echo Dot* were also considered to carry the potential for acquiring vocabulary and meaningful communication according to the views of language learners (Dizon & Tang, 2020).

ELLs in this study also reported enhanced listening skills after practicing English through dialogues with an AI chatbot with a native American accent. The findings also suggested increased confidence levels in their listening skills after practicing linguistic skills with the AI chatbot. Supporting the views of the ELLs in this study as to the contribution of AI chatbots to the development of listening skills, studies reported gains in the listening abilities of ELLs.

For instance, in a study, the acquisition of past -ed allomorph by ESL learners after using an AI chatbot has also been explored to demonstrate if it would help enhance the awareness of phonology, perceptions, and production of English language learners. According to the findings, there was an improvement in their scores in terms of phonological awareness after the implementation. However, there were no significant differences regarding perception and production. The scores of learnability, usability, and motivation were also high, suggesting their willingness to use AI chatbots for language practice. Thus, the study provided evidence for the connection between perception development through listening to AI chatbots and the production of L2 morpho-phonological output (Moussalli & Cardoso, 2021).

Using AI chatbots also enhanced critical thinking skills and language learning satisfaction and increased instances of interaction (Belda-Medina & Calvo-Ferrer, 2022). For instance, after getting engaged in a chat with a chatbot before a group discussion, the learners' productivity and critical thinking skills were enhanced (Goda et al., 2014). The use of chatbots also helped to develop opportunities for collaborative learning and cognitive engagement through the use of AI chatbots in online discussion groups (Hew et al., 2022; Kuhail et al., 2023).

Another subject asserted regarding the pedagogical opportunities of AI chatbots for English language learning was the cultural knowledge enhancement by ELLs in this study. According to their beliefs, the AI chatbot possessed knowledge about various cultures. When ELLs interact with the AI chatbot, they can learn about other cultures and their own cultures. However, they still believed that this cultural knowledge aspect of the AI chatbot could be enhanced more through the application's connection to the internet, especially when they felt that the AI chatbot's answers were superficial regarding the knowledge of specific cultures. However, all in all, AI chatbots were considered beneficial tools to be employed in learning about a variety of cultures while practicing the skills in the English language.

Supporting the participants' views in this study, using AI chatbots has also been argued to be an influential factor contributing to culture-integrated language learning in other studies (Mageira et al., 2022). Through AI chatbot use, learners could also be exposed to cultural components that would add to their knowledge of the target culture and cultural awareness. Thus, their intercultural competence would develop, allowing them to communicate in the target culture's language (Divekar et al., 2021). Thus, chatbots could encourage learners to search for and incorporate new knowledge (Chuah & Kabilan, 2021).

The effect of AI chatbots on factors such as motivation, anxiety, and willingness to communicate (H.H.J. Chen, Yang & Lai, 2023; Jeon, 2022a; Tai & Chen, 2020) has been investigated. Regarding the beliefs of ELLs about the effects of AI chatbots on learning languages, language learners explained how the use of chatbots impacted their

stance on English lessons. Most of them stated that employing chatbots caused them to have a positive approach towards English classes. A learning configuration that allowed them to practice English language skills was effective because it helped them to feel relaxed, as they did not have to think about other interlocutors. They did not fear making mistakes or being judged, as in the case of the presence of human partners. So, it helped to reduce their anxiety about speaking (Jeon, 2022b). The studies also suggested that interaction with a chatbot helped learners construct ideas and participate in classroom debates. It also helped to create a stress-free communication environment by reducing the anxiety of language learners (Bailey et al., 2021; Guo et al., 2023). When the anxiety level of learners was low, they would have a good level of communicative competence, which suggested an L2 WTC that would result in improved oral communication skills and L2 learning (Tai & Chen, 2020a).

In this study, ELLs mentioned that their confidence in speaking English increased after conversing with an AI chatbot. For instance, some learners mentioned that although they might not have advanced linguistic skills in a foreign language, interacting with an AI chatbot supported their self-efficacy beliefs about communicating in a foreign language because they were able to have the chance to practice their language skills and observe that they were capable of having a communication using their knowledge about the target language and practicing their linguistic skills. Corroborating the findings of this study, studies conducted so far have discovered that the use of AI chatbots increases confidence in speaking in a second/foreign language (H.H.J. Chen, Yang & Lai, 2023; Jeon, 2022b; N.-Y. Kim, 2018b; H.S. Kim, Kim, & Cha, 2021; S. Lee et al., 2011; Tai & Chen, 2020b; H. Yang et al., 2022). Studies also reported significantly higher self-efficacy in the experimental groups using AI chatbots (N.-Y. Kim, 2018b); thus, the use of chatbots was put forward as a beneficial learning activity having the potential to enhance the confidence and learning enjoyment of language learners (Tanghe, 2009).

For instance, in such a study, using the *Alelo Enskill* platform, language learners were able to get personalized feedback from AI avatars through analysis of learner speech and error correction. At the end of the study, language learners developed their fluency and enhanced self-confidence (Johnson, 2019). In several other studies, the use of AI

chatbots has also been reported to promote interest in learning languages (Fryer & Carpenter, 2006; N.-Y. Kim, 2018b; H.S. Kim, Kim, & Cha, 2021; S. Lee et al., 2011).

Research thus indicated that learners using AI chatbots demonstrated increased levels of learning interest due to the affective feedback from chatbots (Han, 2020b; N.-Y. Kim, 2018b). In line with the findings in the previous literature, ELLs in this study also reported that using an AI chatbot motivated them to continue practicing English language skills. When their self-efficacy beliefs showed development after the intervention, their motivation to speak in the target language also increased, as demonstrated by the opinions of language learners in the qualitative findings. According to researchers, implementing AI in language learning helps motivate learners to engage in communicative activities and stimulate or elevate interest in learning languages (Diwanji et al., 2018; Fryer et al., 2019). Supporting the findings of this study, the use of AI chatbots enhanced motivation for learning in other studies employing AI chatbots as well (Chien et al., 2022; Han, 2020; N.-Y. Kim, 2018b; S. Lee et al., 2011; H. Yang et al., 2022). These studies demonstrated that AI chatbots could increase the motivation of language learners for speaking (Han, 2020), listening practice (Chien et al., 2022), and vocabulary learning (N.-Y. Kim, 2018b). The motivation levels of learners could increase thanks to the use of AI chatbots and their contribution to social presence, attention, and emotional attachment (Ebadi & Amini, 2022). AI chatbots could promote language learners' engagement, develop language skills, and maintain motivation (Kohnke, 2023; Smutny & Schreiberova, 2020). According to the findings of a meta-analysis study, AI chatbots also significantly affect motivation and interest in language learning, self-efficacy levels of learners, perceived learning value, and anxiety. Therefore, the findings of these studies contribute to the idea that the use of AI in the practice of second or foreign language skills also predicts an increase in motivation and interest in learning languages.

Pre-service teachers in other studies also found chatbots effective for engaging in learning through an interactive language learning experience (Sung, 2020). The results of the previous studies demonstrated that students considered AI chatbots to be engaging and influential teaching tools to be employed in speaking classes. According to students, AI chatbots would support students and teachers because they effectively

learn to speak, foster critical thinking, and enhance language learning enjoyment. The students also stated they would use them in their future teaching careers, especially in speaking classes (Hapsari & Wu, 2022).

According to learners' opinions in many studies, AI chatbots support language practice activities in a fun, effective, and motivating way. It provides opportunities to practice the language authentically, providing meaningful interaction, scaffolding, and immediate feedback (Tai & Chen, 2020b). In a study conducted to explore perceived AI presences, it was also discovered that cognitive presence significantly predicted language enjoyment, which could mean that thanks to the support learners received from the AI chatbot through factors such as getting feedback on pronunciation and having encouragement, learners might construct a positive attitude towards learning English (X. Wang et al., 2022). Similarly, corroborating the findings of this study, in Dizon and Tang (2020), the most recurring theme in interviews with language learners was learning enjoyment, and many learners stated they had fun while talking to *Alexa*. In these studies, learners positively perceived AI bots' use in language learning due to having an enjoyable and motivating experience (Underwood, 2017).

However, in contrast to Tai & Chen (2020a) and X. Wang et al. (2022), the lack of feedback on pronunciation and intonation of the words was a concern for the ELLs in this study. ELLs also expressed a need to get help from the AI chatbot when they cannot develop the proper words in a conversation. Thus, learners asked for explicit feedback from the AI chatbot on word use, pronunciation, and intonation. Therefore, the need for error correction by the AI chatbot was a feature suggested to be considered by the application developers.

The findings from the qualitative data in this study also indicated that ELLs suggested some improvements in the efficiency of AI chatbots for pedagogic use of the AI bots. The first suggestion was the incorporation of gamification elements. By including gaming features such as a guessing game, language learners would benefit from using the application with enhanced motivation to practice their language skills. Another idea for developing the interface of the AI chatbot was the inclusion of synonyms on screen. ELLs argued that reaching synonyms would help users instantly check for

unknown words and help the vocabulary development of language learners. Another suggestion by the ELLs concerned the need for a vocabulary learning section that keeps track of the words used by ELLs and the words used by the AI chatbot. Thus, learners could keep track of the words employed throughout the weeks through a notepad. Another pedagogical implication that emerged in the qualitative data was integrating the informal use of language, idioms, and sayings into the speech database of the AI chatbot. Moreover, adding a section for grammar practice for language learners was another suggestion by the learners. Also, when learners were interacting with the AI chatbot, it sometimes gave repetitive answers while talking about cultural subjects. Therefore, incorporating knowledge about the cultural elements emerged as another need for developing the AI chatbot *Replika* in this study.

Therefore, using chatbots has led to neutral outcomes in this study. There were positive and negative learner experiences during the implementation of the AI chatbot by learners in this study, and they put forward many suggestions regarding improvements for pedagogical benefits. However, in the literature, it was also the case that the use of chatbots has also been found to demonstrate unsatisfactory results regarding language learning outcomes, and it was attributed to several factors. For instance, according to Coniam (2014), chatbots cannot carry out a goal-oriented conversation, which is crucial for learning languages. It was also discovered that no significant differences were found favoring the experimental group concerning engagement in learning (C.C. Liu et al., 2022). It was also put forward that using chatbots would exert extra cognitive load on learners (Huang et al., 2022).

Moreover, according to a meta-analysis, no significant impact of the AI bots on learning outcomes was found, while an utmost effect on production activities and knowledge tests was discovered (Bibauw, François & Desmet, 2022). However, in contrast to the findings of these studies reporting the adverse outcomes, there were also wider-scale studies demonstrating the impact of AI chatbots on learners. In such a study, the findings of a meta-analysis showed that AI chatbots had a significant effect on students' learning outcomes in the areas of performance in language skills. There were also the moderating impacts of the level of education and educational background. That is, higher education learners benefited more from using AI chatbots,

and shorter interventions were more influential in improving learning outcomes (R. Wu & Yu, 2023). In another study, a meta-analysis study investigating the impact of chatbot-assisted language learning demonstrated that using chatbots had a positive and moderating impact on learning outcomes. Moreover, supporting the findings of R. Wu and Yu (2023), chatbots were more beneficial for students at K12 levels rather than tertiary level students (S. Zhang, Shan & Lee et al., 2023).

Also, there were significant benefits in vocabulary, listening, speaking, and other aspects. The researchers also found that using self-created chatbots was more effective than present chatbots aimed at general audiences. It was because these chatbots were developed in line with the experiment's objectives, and additional utilities like instant feedback and dynamic assessment were included. Additionally, speech-to-text chatbots were more beneficial than text-based chatbots in language learning, allowing learners to practice language skills more actively and naturally (S. Zhang, Shan & Lee et al., 2023).

Using more basic answers for beginner-level learners was another suggestion by the ELLs in this study. However, there were conflicting views regarding this issue. For instance, in this issue, another learner stated that this application was suitable for the use of learners with lower proficiency levels. Therefore, there were neutral beliefs about how much language learning enjoyment was provided to particular learners in this study.

Regarding the satisfaction level of students according to their proficiency levels, it was discovered in Yin and Satar (2020) that language proficiency level was a determining factor in the satisfaction state of the learners. According to researchers, learners with lower language levels would find chatting with chatbots more productive and gain more from these conversations with the AI chatbot. In contrast, higher proficiency level language learners express dissatisfaction with using chatbots to practice a language (Yin & Satar, 2020). Moreover, it was argued that using AI chatbots was more efficient for learners with lower language proficiency. In comparison, learners with higher proficiency would regard AI chatbots as less enjoyable and useless (Yin & Satar, 2020).

Nevertheless, although positive and neutral experiences have been reported regarding the pedagogical uses of AI chatbots for learning and practicing languages, they have valuable promises for the future of technology-assisted language learning. AI chatbots encourage learners to have meaningful practice in the target language and reduce language learning anxiety. Furthermore, autonomy could be elevated by using AI chatbots as both in-class and extracurricular learning tools (Dizon et al., 2022; Katsarou et al., 2023).

The findings of this study suggested various pedagogical benefits of AI chatbots offered to language learners. Despite some of the drawbacks learners might encounter using these tools, they have the prospect of supporting language learners by providing practice in the target language. Thus, through voice-enabled AI chatbots, opportunities for language practice, motivation, and confidence are elevated, eventually contributing to the development of language abilities (N.-Y. Kim, 2016). Conducted as an out-of-class extracurricular activity, this study also has significant potential for enhancing autonomous technology-assisted language learning behaviors of ELLs in this study, as they had chances to get engaged in AI-supported interactive tasks, which could further elevate their interest in using similar other technologies for foreign language learning in the future.

5.6. Discussion in Relation to the Social Affordances of Using AI Chatbot Technology for English Language Learning

In an EFL classroom, factors such as speaking anxiety could negatively impact the learners (Hapsari & Wu, 2022). Researchers in computer-assisted language learning have been studying the psychological impacts of technology on learners, as the analysis of psychological conditions in cases learners utilize technological applications reveals the optimal circumstances for learning (Dörnyei, 2005). In these circumstances, chatbots are suggested in tasks, as they provide psychological benefits for language learners. Conversing with a chatbot helps create a stress-free language learning environment for learners, which decreases students' social anxiety stemming from peer pressure. Consequently, learners can talk in English more freely and participate dynamically in learning activities (Hill et al., 2015; Jeon, 2022b).

Moreover, creating a comfort zone by the AI chatbot also led to a comforting, friendly language practice environment for the participants. According to the ELLs in this study, the AI chatbot *Replika* responded in a friendly manner. They stated that they were able to form a close personal relationship with *Replika* because it responded in a comforting, friendly, and understanding tone. This finding was unsurprising because *Replika* was a chatbot designed to function as a therapeutic AI chatbot. ELLs also stated that they felt that the experience of talking to an AI chatbot was comforting, as the AI did not contradict their opinions but supported them or gave its own opinions on the issue. Another point concerned the feeling of being more at ease talking to a robot than a human being. ELLs believed that even if they made mistakes while talking to the AI chatbot, they did not feel the anxiety they had while talking to real people. For this reason, they could talk freely without the concern of being evaluated or judged on their linguistic abilities in the language. One further characteristic attributed to the AI chatbot was patience. ELLs mentioned that no matter how long they spoke, the AI listened to their speech, which allowed learners to practice the target language, especially in an EFL setting where they need a language-speaking partner.

According to the affective filter hypothesis, learners could be affected by numerous outside factors in terms of motivation, anxiety, and self-confidence. This, in turn, would lead to difficulties in learning a language. Therefore, interacting with a chatbot would help lower the affective filter of language learners. When learners successfully interact with a chatbot, they carry this experience into everyday interactions with real people. That is, they would be able to feel the same confidence they felt talking to a chatbot while interacting with real language partners (Krashen, 1981; Tanghe, 2009). Teachers could implement AI chatbots to help students decrease their speaking anxiety, making learners less stressed when talking in front of a group of people (Hapsari & Wu, 2022).

Other studies also found that using AI chatbots could alleviate the anxiety levels of learners (Han, 2020). In a study conducted with language teachers in a Korean EFL setting, teachers' opinions regarding the usefulness of AI chatbots for language learning were demonstrated. They believe AI chatbots provide learners with safe spaces for language practice (Kent, 2021). Through AI chatbots, shy and worried

learners could practice their language skills in a relaxed learning space without teachers or peers (H.L. Hsu, Chen & Todd, 2023; Jeon, 2022b). Learners had positive opinions about using chatbots, and they enthusiastically interacted with them. They felt more relaxed and comfortable with a chatbot partner than a human partner (N.-Y. Kim, Cha, & Kim, 2021). Thus, the affective barriers to learning decreased, as they did not fear losing face while speaking a foreign or second language (Fryer & Carpenter, 2006b; Tai & Chen, 2020a; Underwood, 2017).

In another study, language learners considered the AI chatbot *Alexa* as a language partner with which they could freely chat without fear of negative evaluation. Therefore, learners were more relaxed talking to *Alexa* than to human partners. The communication was thus learner-centered, with lowered self-consciousness and enhanced communicative confidence (Tai & Chen, 2020a). Thus, the findings suggested that learners who feel less anxiety about speaking in a foreign language using AI chatbots can interact more effectively and without errors.

The findings of another study demonstrated that the anxiety and stress levels of language learners with higher proficiency chatting with both AI chatbot *Replika* and NSs synchronously declined. Nevertheless, there was no change of beliefs regarding their motivation. However, for lower-level learners, there were improvements in their motivation levels, as well as an experience of reduced anxiety and stress after interaction with the AI avatar (Kang, 2022).

It is also argued that using AI chatbots also creates a social presence among language learners because chatbots can simulate human-resembling social communications, which helps create a relaxed environment for learners (Chuah & Kabilan, 2021). In line with this argument, anthropomorphism appeared as another theme regarding the social dimensions in this study. Some ELLs in this study regarded the AI chatbot as human-like. According to their beliefs, the AI was like a natural person, capable of doing more than expected from a robot.

According to the views of some of the ELLs, the AI learned from its interactions with the learners each week and collected more information about the learners; therefore,

the conversation felt more natural after learners had used the application for some time. Moreover, some ELLs stated that the AI chatbot could sometimes apologize to learners; for instance, when the chatbot interrupted a learner's speech, it apologized for the intrusion, another social skill the AI chatbot possessed. Other personal attributes of the chatbot were being polite, respectful, kind, chatty, easy to talk to, and fun. ELLs thought *Replika* had a kind attitude toward them in their conversations and always respected their opinions. ELLs also enjoyed their conversations with the chatbot, as they stated that the AI chatbot could make jokes and that its statements were humorous.

According to Katsarou et al. (2023), the bots build realistic settings for human-machine interactions, adjusting to the needs of language learners. Through a more stress-free environment, language learners are more willing to participate in conversations with enhanced learning outcomes, enjoyment, and confidence.

Corroborating this study's findings, in cases considered realistic speaking partners, chatbots increased students' motivation to employ chatbots for learning languages (Jeon, 2022a). Similarly, in another study, Timpe-Laughlin et al. (2022) discovered that when teachers considered chatbots authentic speakers, they were more willing to use them in their classes. However, when students considered chatbots to be machines, it caused them to be reluctant to talk to them in English. Thus, these students preferred to talk to human partners like their peers. Therefore, it is necessary to make chatbots as authentic as possible, making them more feasible for pedagogical use (Jeon, 2022a).

On the other hand, ELLs in this study also stated that the AI chatbot had the characteristics of a machine that did not resemble a real conversation partner. First, the voice of the AI chatbot was found to be robotic by the learners. According to the learners, the AI chatbot had a robotic voice. Therefore, they expressed a need to incorporate a more humane resembling voice to the bot. Another social constraint, as expressed by ELLs in this study, was the lack of emotions demonstrated by the chatbot, as the AI chatbot in this study was incapable of having feelings and holding a natural conversation. The ELLs also stated a need for the AI chatbot to be friendlier. Thus, some learners believed that the AI bot could not bond with its users in a friendly

manner. The ELLs in this study also mentioned that the AI chatbot had a robot-like attitude. Therefore, they indicated they preferred talking to their human peers instead of a robot like an AI chatbot.

Consequently, for the learners considering the AI chatbot too robotic, the motivation to use it for language practice could have been higher in this study. They did not demonstrate a willingness to use AI chatbots for their future learning. The findings of the current study are consistent with those of Chuah and Kabilan (2021), Jeon (2022b), and Zou et al. (2020), demonstrating both positive and negative learning experiences in the use of an AI chatbot. For instance, in Zou et al. (2020), language learners' motivation levels got lower because of a similar belief that chatbots cannot replace human partners.

Similarly, according to the results of Jeon (2022a), when the AI chatbot was considered a machine-like conversational partner, language learners' motivation to use the application decreased. In the case of perceiving the chatbot as a calmer speaking partner, it impacted users' perceptions and increased their motivation to use it continuously. Thus, some ELLs in the study did not like practicing English language skills with chatbots because they preferred human partners with whom they could collaborate. They also preferred to study the language face-to-face with their peers rather than chatting with a chatbot individually. Thus, they had low motivation to speak in English with chatbots. It has also been argued that AI chatbots need improvement regarding artificial intelligence, as teachers and ELLs need to develop their responses, which teachers and ELLs consider mechanic, unnatural, and deprived of social presence (Chuah & Kabilan, 2021).

Another social constraint put forward by the participants in this study was the lack of human-like avatars. Namely, ELLs in this study did not consider the physical appearance of the AI chatbot *Replika* as realistic as it should have been. Therefore, some learners stated a need to change the avatar's appearance. It was also discussed that the avatar was always the same gender. Although the avatar's gender could be changed from the application's settings, learners always had an avatar with the same

gender due to sharing the same account to log in to the AI chatbot application with their classmates.

Apart from its physical appearance, there was a lack of use of gestures by the AI chatbot. According to the opinions of ELLs, rather than visually demonstrating bodily gestures, the chatbot indicated gestures through subtitles or voice commands, which did not show the characteristics of a natural conversation. The AI chatbot also had deficiencies in the use of facial expressions. Sometimes, it did not use any mimics, or even when it used mimics, the learners expressed that the demonstration of mimics had better be developed. Thus, learners preferred that this area be developed in the design of the future generation of AI chatbots.

ELLs in this study also mentioned that the AI chatbot had no personality. Learners stated that each time they started a new conversation, the chatbot behaved as if they were speaking for the first time. Thus, there were also cases when the AI chatbot did not suffice to meet the users' expectations in terms of forming a social bond through getting to know each other gradually during application use each week. This situation caused ELLs to feel emotionally distant from the AI chatbot and they suggested developing *Replika's* personality characteristics.

Another issue voiced by the ELLs in this study was the lack of social skills by the AI chatbot. The AI chatbot was considered deficient in social aspects. Learners stated that it was not able to comprehend abstract concepts. ELLs also mentioned that the AI chatbot did not apologize for intruding on the learners' speech. The AI chatbot also could not understand some of the learners' jokes and did not make them often. It was defined as being deprived of empathetic skills by some learners, too. In line with the findings of this study, negative experiences after using AI chatbots have also been reported in El Shazly (2021). The use of AI chatbots has been investigated in an EFL setting in Egypt with undergraduate language learners. According to the results, the anxiety levels of learners did not decline after interacting with an AI chatbot. Moreover, the decline in the motivation levels of learners was attributed to the nature of human-chatbot interaction. It was therefore suggested that whereas during a natural interaction, teachers adapt to the topic-shifting needs of the learners through the

observation of non-linguistic hints by learners, machines could not suffice to meet the requirements of social interaction, as they lacked the characteristics such as complex reasoning, empathy, etc.

One issue that ELLs frequently repeated in the interview and reflective journals was that the AI chatbot had problems regarding the formality of the speech, as it talked to the learners in a personal tone that was too personal at times, according to the opinions of some of the ELLs in this study. That is, the tone of the speech was offensive to some of the users. Although this situation was controllable by adjusting the settings of the application, learners in this study still had problems regarding the formality of speech. ELLs also stated that the AI chatbot mentioned delicate subjects such as religious beliefs, which they believed might not be pedagogically suitable for class use. Moreover, as this chatbot was created for therapeutic purposes, it often diverged the topics to emotional subjects, which was found depressing by some of the ELLs in this study.

The novelty effect was a factor that was argued to affect the experiences of language learners negatively. It was pointed out that due to a novelty factor, language learners could lose their initial interest in chatbots for learning languages after some time (Fryer, 2019; Fryer et al., 2017). Initially, a language learner might show interest in using chatbots due to the newness of the experience. However, this interest might drop during the implementation after this initial interest has passed away (Fryer et al., 2017). It was also argued that chatbots could cause learners to have novelty impacts and feel adverse emotions that would thus lead to unsuccessful attempts at learning.

For instance, in a study, interest levels decreased in chatbot-human groups due to the novelty effect and non-authentic use of the language after using the AI chatbot for 12 weeks. In contrast, in human-human chat groups, there was no decrease in the interest level (L. K. Fryer et al., 2017) in contrary to the previous literature stating that the exploitation of chatbots would benefit learners in terms of motivational factors (Fryer & Carpenter, 2006b; Hill et al., 2015; Weizenbaum, 1966). Supporting this belief, it was found in a meta-analysis study that when the intervention duration was shorter,

that is, when learners experienced using AI chatbots for a shorter period, it had a more significant impact on learning outcomes (R. Wu & Yu, 2023). Thus, it was argued that learners could be even more eager to use AI chatbots when the intervention time is shorter. It thus might improve their motivation and interest in language learning and performance in the language for a temporary time (Fryer et al., 2017). As soon as the novelty effect subsides, it could be challenging for learners to maintain their interest and motivation and to have a good learning performance (Haristiani, 2019).

However, the literature has conflicting evidence regarding an ideal intervention duration. For example, according to S. Zhang, Shan & Lee et al. (2023), regarding instruction duration, when the implementation period was between 1 to 3 months, chatbots were discovered to be more influential for learners. On the other hand, it was the least effective when the study lasted for a shorter time frame than a month. This study was also carried out within two months. Learners might have demonstrated an initial interest in using the AI chatbot, which might support the presence of negative comments regarding the technical, pedagogical, and social affordances of using AI chatbots for practicing languages while using the application. However, the duration of two months was also an ideal period, as supported by the earlier evidence by S. Zhang, Shan & Lee et al. (2023) and R. Wu and Yu (2023). For this reason, this issue needs further investigation through longitudinal data collection. Therefore, although there were no statistically significant results in every language skill being evaluated after the intervention, the results from the qualitative data analysis demonstrated that there were gains in vocabulary and grammar knowledge enhancement, listening and speaking skills development, and increased self-efficacy regarding communicative confidence together with reduced foreign language speaking anxiety beliefs of ELLs in this study.

5.7. A Framework for AI Chatbot Use for EFL Teachers in Teaching Target Skills Based on the Findings

In this section, a framework consisting of the components of optimum qualities of AI-based chatbots, suggestions of use by language teachers in terms of technological, pedagogical, and social aspects, a proposed conceptual model for chatbot-assisted

language learning, and a module for language teacher training on AI-based chatbots are proposed.

The findings of the present study collected from listening and speaking language proficiency exams, interviews with English language learners, learner diaries, and AI-based chatbot-language learner interaction data have been merged towards the aim of presenting the outcomes for this study for EFL classes, language learners, teachers of foreign or second languages, teacher educators, application designers, and future researchers within this field.

5.7.1. Optimum Features of AI-based Chatbots for Learning and Practicing English as a Foreign Language

Due to the developments in text-to-speech engines and open-access artificial intelligence ChatGPT platforms, artificial intelligence-based chatbots using an aural mode of communication have become more widespread in recent years. Their use has expanded to many areas, including health care, finance, and educational systems. Within the field of education, AI-based chatbots have been used both as student support systems by responding to frequently asked questions by college students and as pedagogical tools through the integration of AI chatbots both as in-class and out-of-class learning tools. Namely, AI-based chatbots have been used in English language teaching to gain grammar and vocabulary knowledge and practice writing, reading, listening, and speaking skills. Thus, due to their increased implementation in the language teaching field, it has become more vital to answer the question, “How should an ideal AI chatbot that would interact with EFL learners be like?”

AI-grounded chatbots, also called contextual chatbots, utilize NLP and ML to comprehend the user’s intent, respond accordingly, and constantly learn from user interactions (Smutny & Schreiberova, 2020). Owing to their advanced conversational systems, the AI tools can connect with their users simultaneously through a text-based or voice-enabled system. For this study, the focus was on the use of voice-based AI chatbots as language conversation partners. Based on the findings of this study, it could be argued that through the demonstration of language learners’ views and experiences

using interviews, reflective diaries, and AI chatbot-language learner interaction data, and the outcomes for learning through language proficiency tests, a clear-cut picture of optimum characteristics deemed necessary for an AI-based voice-enabled chatbot has been drawn.

Table 54 Optimum technological AI chatbot features based on the findings of this study

Technological	Pedagogical	Social
a. A fully functioning ASR system	a. Helps with unknown words through an info button	a. Human-like characteristics
b. A vast knowledge storage base	b. An omniscient language practice partner	b. Behaves like a real person
c. Varied accent options	c. A feature of explicit error correction	c. Helps construct a comfort zone for learners
d. Moderate rate of speech	d. Use of idioms and sayings	d. Acts like a friend
e. Turn-taking capabilities	e. Possesses information about the target cultures	e. Open to cultural varieties
f. Initiation of interaction actively	f. Has gamification elements	f. Employs a formal tone of speech without mentioning delicate subjects
g. Expanding on previous conversation	g. Adaptable for the use of different proficiency levels	g. Has social skills, i.e. apologizing
h. On-topic utterances	h. Motivates learners to speak in the target language (TL)	h. Has emotional intelligence

Table 54 Optimum technological AI chatbot features based on the findings of this study (*continued*)

Technological	Pedagogical	Social
i. Teachable and getting smarter	i. Incorporates a section on practicing the grammar of the target language	i. Use of realistic gestures and mimics
j. Quick reaction span	j. Offers a personalized language learning experience through adjusting the difficulty level of the interaction by recalling the proficiency levels of ELLs	j. Human-like avatars
k. Remembering past conversation history	k. Allows the practice of listening skills through providing comprehensible input	k. Has a personality, i.e. fun, chatty, sincere, understanding etc.
l. Competent in language skills	l. Voice-enabled AI chatbots offer language exposure by providing a speaking partner	l. Possesses a realistic tone of voice with a natural pace of speech
m. Ability to propose counter arguments	m. Provides negotiation of meaning opportunities by the use of interactive strategies	
n. Having multimedia elements	n. Integration of the use of everyday language	
o. A group talking feature	o. Multiple languages support	

Table 54 above demonstrates the optimum characteristics of an AI chatbot derived based on the outcomes for learning and language learning experiences of English language learners in this study. The characteristics that are expected to be present in an ideal AI chatbot for language learning are represented from the angle of three affordances: (1) *technological*, (2) *pedagogical*, and (3) *social*.

The most prominent codes in the qualitative data have been categorized under the dimensions of technological, pedagogical, and social elements to consider the optimum properties of an AI chatbot for learning languages. Using the analysis results of the qualitative data, first, the technical aspects that could act as a guide for the creation of voice-enabled AI chatbots for practicing languages are demonstrated. Then, the pedagogical characteristics that distinguish these chatbots from those created for a general audience are listed. In the third category, social skills expected from an AI chatbot targeting language learners are demonstrated in a bulleted list. Following is a practical list of suggestions for optimum technical features of an AI chatbot, which are derived based on the findings of this study.

A. Technological aspects

- a. An AI chatbot constructed through the ChatGPT platform needs to recognize the nuances of accents in the speech of English as a foreign or second language and comprehend learner input effectively.
- b. Integrating multimedia elements appealing to language learners with varied learning styles is another essential characteristic to consider.
- c. The AI chatbot needs to wait for an optimum amount of time for the student's turn to end by altering the responding time.
- d. The intelligibility of the non-native speakers' speech by the AI chatbot had better be ideal.
- e. The AI bot should be able to respond to users' queries with on-topic responses.

- f. A fully functioning AI chatbot is to have an ideal reaction span without long pauses between question-and-answer sequences.
- g. The AI chatbot's use of correct intonation and pronunciation in speech is also a necessary property, as these conversational agents need to act as role models to English language learners.
- h. The application users should be able to select the accent of the AI avatar from various accent options available to them.
- i. A conversation turn-taking button that helps the users keep turns in conversation could function well to avoid interruptions by the AI chatbot during speech.
- j. The AI chatbot needs to be capable of having long turns in speech by delivering a lengthy amount of speech in conversation turns.
- k. An effective AI chatbot for language learners would be able to initiate a conversation by asking learners questions based on the learning objectives of a language-specific lesson.
- l. The speech rate of an ideal AI chatbot for language learning needs to be adaptable with slow, medium, and fast options.
- m. A group talking option would also make the AI chatbot adaptable for group discussions.
- n. By specifying the language proficiency of the users, the system could provide learning materials to users according to their language skill level.
- o. It should be equipped with developed augmented reality tools for characteristics such as a realistic avatar with better jests, mimics, and gestures.

- p. A pronunciation help chat screen needs to be incorporated into the AI chatbot for the reference of the language learners whenever needed.
- q. The AI chatbot should be able to connect to the internet instantly during a conversation to check for unknown information.
- r. The subtitles displaying the speech of the AI chatbot need to be of better quality in terms of font type and size and their place on the screen.
- s. An effective AI chatbot functions well on Android and IOS devices, allowing dissemination of the application to a broader audience.
- t. The AI chatbot should be capable of remembering the past conversation history of EFL learners.
- u. The AI chatbot could negotiate meaning with its users by asking for clarification or repeating messages when the user's speech is unrecognizable to its ASR system.
- v. An ideal AI chatbot has a comprehensive knowledge of many topics due to a vast knowledge base in its system that allows it to chat with its users through discussion of various subjects.
- w. The capability to expand on the ongoing conversation by making additional comments, asking for further information on the subject, and asking related questions is another expected technological quality from a well-functioning AI chatbot.
- x. Even the AI chatbot, *Replika*, used in this study has developed since the implementation of the application in 2020 by the English language learners in this study. At the time of the study, *Replika* was utilizing GPT2 and GPT3, yet as of that period, AI chatbots have been getting more brilliant thanks to the most recent versions of ChatGPT, and they

are teachable even more with the practice of the users with the application.

The following is also a list of characteristics an AI chatbot created for the pedagogical use of language learners needs to possess. There are many pedagogical facets to consider while designing an AI chatbot aimed at implementation by learners of foreign languages.

B. Pedagogical Aspects

- a. An ideal AI-based voice-enabled chatbot created for language learners allows users to practice speaking skills by acting as an omnipresent conversational companion.
- b. Aiming at the development of language skills of learners, an AI chatbot needs to help its users practice language skills in task-based conversation scenarios. As the present study's findings suggested, a chatbot created for a general audience might not always fully function in responding to students' queries on specific topics.
- c. Goal-oriented AI chatbots would best meet the learning objectives of a foreign language class, as through the construction of scenarios, i.e., at the cinema, etc., language learners could practice target language structures in a more pedagogical mode. The current AI chatbots could better provide the necessary conditions for conversational practice, unlike task-oriented chatbots created to practice languages.
- d. The current commercial chatbots available for language learning, i.e., Mondly, Duolingo, and Andy chatbots, provide conversational practice based on closed-ended conversation prompts.
- e. That is, the users are given multiple choices in response to verbal input by the chatbot. The users are merely required to choose an option and

read it aloud. Feedback is provided on articulating individual sounds or using intonation in sentences, yet language learners can only practice language skills through structured dialogues. Thus, an ideal chatbot equipped with AI and ASR systems would enable unrestricted dialogue options on various subjects.

- f. Alelo's *Enskill English* (Johnson, 2019) and Sydorenko et al.'s (2019) *HALEF* are some examples of voice-enabled AI chatbots created for English language learners. However, these chatbots are not publicly available, as the researchers developing these AI tools do not share these applications as commercial or freely available gadgets. *Replika* also does not allow its users to chat in a voice-call mode unless registered with a premium account. Therefore, there is a need to design and disseminate publicly available pedagogical AI chatbots for language learners and teachers.
- g. The use of explicit error correction is another optimal quality for AI chatbots. As language learners in this study stated, explicit correction of errors regarding accurate word use, grammar, pronunciation, and intonation would benefit language learners. Through constant feedback, progression toward the language learning goals could be boosted more at an anticipated rate.
- h. An adjustment to different proficiency levels is also a required aspect expected to be present in an ideal AI chatbot for language learners. The chatbot should respond to learners' needs from various proficiencies.
- i. According to the participants' views in this study, a learning track algorithm to track an individual language learner's progress and supply a personalized learning experience is also an expected element. A self-management system is thus necessary for following learners' progress.

- j. Using gamification elements is another element that would add to language learners' language learning enjoyment and motivational states.

According to the students' views in this study, an effective AI chatbot must create a comfort zone by adopting a realistic visual interface. In line with this argument, the subsequent list also outlines the expected social aspects.

C. Social aspects

- a. The AI avatar had better be realistic with its human-like appearance, voice, gestures, and mimics.
- b. It needs to have emotional intelligence to provide comfort to its users through acting like a friend. When language learners feel the presence of support from the AI chatbot, they report feeling relaxed and encouraged to communicate in the target language by lowering the learning anxiety.
- c. Some delicate subjects, such as religious beliefs, politics, and emotional status, were taboo issues that could disturb many users. Hence, application designers should pay attention to this concern for the further development of AI-based chatbots.
- d. Ideally, an AI-based chatbot should have a personality with pre-esteemed birth and cultural origins, set beliefs about certain subjects, and have the capability to interact based on these characteristics. Otherwise, every time learners interact with the AI chatbot, it adopts a different perspective, regarded as self-contradictory and non-humane by several learners.
- e. When students feel the social presence as represented by the views of language learners in this study, they might tend to be more motivated to use this kind of application due to lowered anxiety about speaking in a foreign language and motivation to converse more in the target language.

5.7.2. Guidelines for Teachers Based on the Characteristics of Optimum AI Chatbot Features

Decisions regarding integrating AI-based chatbots into language classes had better be taken considering the profile of language learners. Below are some crucial aspects to consider in the implementation phase of AI chatbots by teachers. Teachers must pay attention to various concerns to obtain maximum benefit for the targeted students from various language proficiency levels. Suggestions regarding using AI chatbots are demonstrated in line with the arguments considered so far. Propositions and concerns for EFL teachers are demonstrated (see Table 55).

Table 55 Suggestions for implementation based on the optimum AI-enabled chatbot features

Propositions	Concerns
a. First and foremost, introduce the characteristics of AI-based chatbots to your language learners.	a. An abrupt use of AI-based tools by language learners without introducing how to implement these tools for communicative language practice
b. Familiarize your learners with AI-based tools through a pre-training phase before implementing them in your classes.	b. Disregarding the needs of your language learners while choosing an appropriate AI chatbot
c. Provide clear guidelines with a list of dialogue prompts, commands, speaker roles in task-based language teaching activities etc. to guide language learners to interact with AI chatbots with maximum benefits.	c. Not matching language learning objectives to the skills practiced by AI chatbots, i.e. textual-based chatbots might not enhance speaking skills as well as voice-enabled types

Table 55 Suggestions for implementation based on the optimum AI-enabled chatbot features (*continued*)

Propositions	Concerns
d. Use both text-based and voice-enabled chatbots according to language skills practiced, i.e. writing essays, reading texts, discussion tasks	d. Even if you are using AI-based chatbots as supplemental tools for your classes, you had better not leave your students without support. Provide assistance whenever needed.
e. Consider the learning styles of your learners while implementing these tools. For instance, using multimodal AI-powered chatbots could match your learning objectives in many cases.	e. The custom-built pedagogical AI agents might not be always freely available to your students, in those cases do not easily lose your motivation and try to seek funding from sponsors or national educational technology development support programs.
f. Create your AI-based chatbot in line with the background of your language learners, lesson units, and national educational policy standards.	f. Developing your own AI-based chatbots would be even better. Do not underestimate the power of these conversational agents for practicing target language skills.

Incorporation of these tools into language teaching programs could be challenging for a variety of reasons. The first issue to consider is the technological infrastructure of your educational setting. AI-based chatbots are equipped with NLP, ASR, and AR technologies, so their implementation requires mobile devices with the latest developments. For instance, if your learners do not have mobile phones with the latest models, i.e., Android, IOS, or other operating systems, you might be unable to implement these tools in your classes. In those situations, consider an alternative application route. You might provide learners with iPads or other digital mobile tools,

or you might consider using the laboratories in your schools, if there are any, as some of these applications are supported for use through a desktop.

Technological readiness also matters in the preparation stage, as you need to consider the technological skills your language learners possess for using AI chatbots. Supporting students before, during, and after phases of the application use is crucial to prevent any issues concerning technical breakdowns and other user-related aspects. Using different AI chatbots might also require different knowledge from the users. For instance, some AI-based chatbots like Apple's *Siri* or Amazon's *Alexa* function based on short commands delivered to AI agents. These agents can mainly provide practice for answering queries regarding specific information, i.e., the weather, the capital city of a country, etc., or carrying out tasks such as playing songs or calling people. Henceforth, in line with the learning objectives, teachers could provide learners with prompts to interact with these tools. Suppose language learners at lower proficiency levels are not supported during these phases. In that case, they might not interact with AI chatbots because they lack the knowledge and skills to direct the conversation skillfully. It is crucial at this stage to help learners achieve learning aims by supporting them with enough guiding prompts for chatbot use.

Regarding the particular characteristics of the language skills practice, different AI chatbot types, such as a text-based chatbot like *Cleverbot* (<https://www.cleverbot.com/>) or multimodal interaction modality chatbots like *Boibot* (<https://www.boibot.com/en/>), and *Eviebot* (<https://www.eviebot.com/en/>) and *Replika AI* could be selected. However, these available chatbots need more qualities listed under optimum features for AI chatbots.

Hereafter, the creation of pedagogical AI chatbots targeting language learners is a crucial step to consider to develop chatbots that would best match the language learning objectives, the national educational policy standards, curriculum goals and objectives, and the specific needs of your language learners with varying learning styles and proficiency levels.

5.7.3. Suggested Conceptual Model in Relation to Language Knowledge Attainment and Skills Practice

Researchers investigating the use of AI-based chatbots in the fields of education and language learning have developed several models offering guidelines to demonstrate the characteristics of chatbots that would be most effective for learning. Based on a review of studies using AI chatbots in a variety of fields for learning in general, R. Zhang, Zou, and Cheng (2023) put forward a “RAISE” model with its components of “repetitiveness, authenticity, interactivity, student-centeredness, and enjoyment” (p. 13). According to the RAISE model, employing AI chatbots provides opportunities for repetitive interaction with the target subject in an authentic learning environment. Using these tools, a student-centered learning environment with the potential of a motivator leading to learning enjoyment would be constructed.

Supporting the framework of R. Zhang, Zou & Cheng (2023) model, Jeon, Lee, and Choe’s (2023) systematic review also suggested a conceptual framework with the factors of “goal-orientation, embodiment, and multimodality” (p. 7) after exploring the use of voice-enabled AI chatbots in various language learning settings. Concerning the first dimension in the framework, Jeon, Lee, and Choe (2023) discovered a trend among the studies reviewed that researchers either developed goal-oriented chatbots targeting language learners or adopted the use of already present non-goal-aligned configurations created for the use of general audiences.

Although they argued that both types had unique benefits, the authors supported goal-oriented AI chatbots in learning and practicing languages. It has been put forward that this goal orientation would offer four affordances for teaching languages: task-based language learning, support during instruction, adaptable exercise difficulty, and customized learning.

The researchers also argued that “interlocutor variables related to embodiment, such as appearance, non-verbal cues, facial expressions, and gesturing, affect how learners experience interaction with other interlocutors” (Jeon, Lee & Choe, 2023, p.10) and thus have an impact on learning outcomes or experiences. Moreover, as Jeon, Lee, and

Choe (2023) indicated, the embodiment of avatars adds to constructing an immersive setting and forming personal attachments. Displaying interaction transcripts on screens through a multimodal mode also helps self-monitoring, correction, and understanding.

Moreover, Topsakal and Topsakal (2022) also suggested a framework for developing AI-based chatbots targeting young learners of foreign languages. The proposed framework integrated augmented reality elements, voice-enabled chatbots, and large language models (LLM) of ChatGPT. In their framework, the elements advised to be included in the design principles of AI chatbots were gamification, a reward system, and the adoption of cognitive, audio-lingual, and communicative language teaching principles, which would be considered suitable factors for the targeted group of young language learners in this case.

Although conceptual models have been proposed in the field of AI chatbot creation based on literature reviews, more research is needed to propose a model for AI-based voice-enabled chatbots based on the empirical findings of studies such as the study at hand. For this purpose, the findings of this study were used to specify problematic issues as well as well-functioning aspects in the use of AI chatbots based on the experiences of language learners and propose a model for the development of a future generation of pedagogical AI chatbots. Therefore, the findings of this quasi-experimental study would yield a contribution to the field through the proposition of a model for developing a voice-enabled AI chatbot for English language skills practice.

A novel approach for I-CALL and I-MALL-assisted language learning, building upon the integration of AI-assisted chatbots in language classes, is suggested in this thesis. Concerning the optimum characteristics an ideal AI chatbot created for language learners should have, a REPOST model is proposed: Reciprocity, Evaluation, Personalization, Omnipresence, Socio-cultural (elements), and Teachableness (see Figure 7).

First, *reciprocity* refers to the initiation of conversation by both parties during an interactional sequence. An analysis of the AI chatbot-English language learner interaction data and qualitative findings has demonstrated that the AI chatbot *Replika*

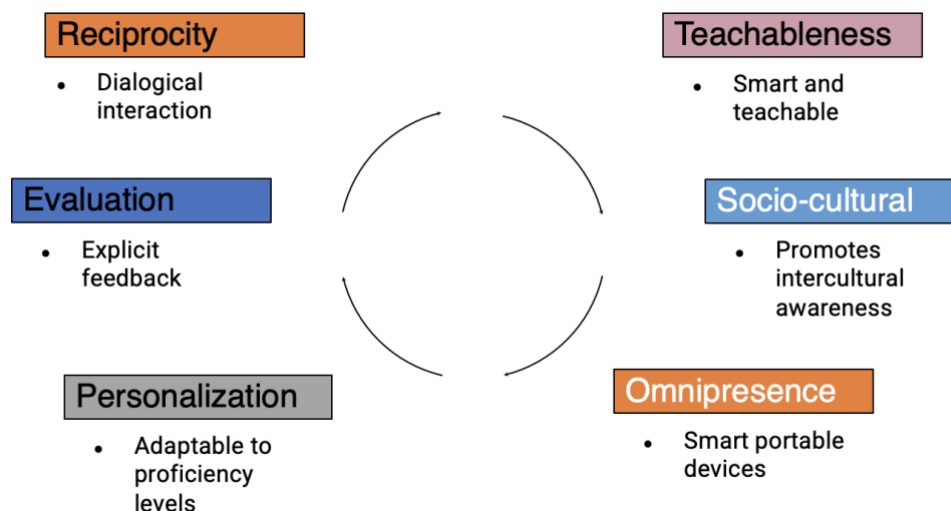


Figure 7 The REPOST Model

did not initiate conversation as much as language learners in this study. Learners expressed in the interviews and their reflective diary notes that they had expected more conversation initiated by the AI chatbot. Contrary to their expectations, learners were mostly conversation starters during their interactions. Therefore, a need for better turn-taking capabilities was a recurrent theme mentioned by the participants in this study. Although the initiation of conversation was helpful for language learners because they could produce a good amount of output in the foreign language, the conversation stopped in some cases, especially when learners could not comment further. For these reasons, an ideal AI chatbot should provide language learners with reciprocal interaction opportunities to promote the utmost motivation to practice language skills in a foreign language. Contrarily, a one-way interaction might only allow a few conversational chances for language learners, as they might feel unmotivated, unsupported, and overlooked, and eventually might stop using these conversational AI agents altogether. Hence, reciprocity is the first component in our proposed conceptual model for AI chatbot creation processes.

Based on a task-based theoretical basis, scenario-based interaction modes in AI chatbots would allow the use and practice of language structures through a reciprocal mode of speech. Through task-based activities, language learners could practice specific language structures in situational conversational contexts such as role-playing

scenarios. Furthermore, the instructional chatbots created for scaffolding target language learners could also deal with communication breakdowns by supporting language learners through interactional strategies such as clarification requests, confirmation, and comprehension checks. Through the use of strategies by AI chatbots like simplifying utterances, codeswitching (Ayedoun et al., 2020), scaffolding by restating and reformulating, etc., AI chatbots could also provide language learners with interactional support.

Evaluation of input from language learners through the provision of explicit feedback is another essential characteristic that has been incorporated into this chatbot development model. The findings of this study demonstrated that through the use of a multimodal chatbot, language learners were exposed to both oral input through interaction with an AI conversational chatbot and textual input using the transcripts of the conversation on screen. Both of these processes permitted implicit self-correction and self-observation of the errors by the learners. On the other hand, besides leading to implicit feedback, the AI chatbot might also need explicit error correction, as expressed by the language learners in this study. An AI-based chatbot created for the pedagogical uses of language learners had better provide feedback on issues such as correct forms of grammar, the choice of vocabulary, and accurate articulation in terms of pronunciation and intonation in the target language. For this purpose, a pronunciation help chat screen and a feature of an info button to help with unknown words were among the suggestions stated by the participants in this study. Thus, the evaluation component in the form of explicit feedback is another element proposed to be included in the development process of a pedagogical AI chatbot.

Another component in our model is personalization, which stands for the learning experience expected to be delivered by AI-based chatbots based on the characteristics of individual language learners, i.e., age, language proficiency level, etc. First, the procedure for this outcome would consist of creating custom-made AI chatbots for language learners to use. Grounded in the experiences of chatbot users in this study, a chatbot first needs to adjust to learners' language knowledge from different proficiency levels. The second feature related to the personalization of language learners' experiences could be realized through the incorporation of necessary technical features

so as for the AI chatbots to remember the user history; namely, based on the success of past conversations, the chatbot could utilize the question prompts that language learners best reply to encourage communication with enhanced learner motivation and interest. Consequently, it could be possible for language learners to have a personalized learning experience thanks to the use of chatbots that could provide tasks according to their proficiency levels, interests, and preferences.

The *omnipresence* feature in this model indicates that AI chatbot technology should offer its users autonomous language learning and practice opportunities. The omnipresence property is put forward as an essential component of this model because the use of chatbots helps learners to practice their target subject as many times as they need without any limitations, and it allows an easy-to-reach practice of the target subject through AI-based learning support systems. Considering the latest developments in technology, the use of AI chatbots for language learning would also become more widespread by incorporating AI chatbot technology into wearable devices such as Google Glass or smartwatches. This way, language learners could have a more mobile companion for language practice, which is present nearby.

Concerning the *socio-cultural* component in this model, one of the first elements to consider before embarking on the creation process would be training the AI chatbot with the inclusion of cultural materials representative of many cultures worldwide. In line with participants' views in this study, including exclusive information about languages, cultures, traditions, and customs of various nations worldwide would be one crucial aspect worth considering in designing a pedagogical AI chatbot targeting language learners. In this way, learners would feel culturally embraced upon seeing that the AI chatbot knows about their traditions and admires their cultural values. Furthermore, conducting an intercultural unit on teaching languages is possible when the AI chatbot is provided with cultural information in its knowledge base. However, as put forward by (Petrović & Jovanović, 2020), although machine learning chat script application interfaces such as Google Dialogflow and Tensorflow could be used to create chatbots, the design process employing data representing a variety of languages is reported to be 'biased,' as the current APIs are mostly predetermined to employ the English language. Thus, machine learning algorithms in these APIs might not be fully

able to learn languages other than English. Nevertheless, by compiling datasets from various countries in the chatbot design process, sociocultural contexts would also be considered with the integration of cultural components in these datasets. The outcomes of this process could be effective, especially for creating AI chatbots that would suffice to meet the expectations and needs of foreign or second language learners.

Our model also suggests that the factor of *teachableness* is an indispensable element that needs greater attention by teachers, application designers, and other stakeholders taking part in an application development process. The chatbots constructed through formerly broadcasted generative models can already learn from conversations with their users by detecting contextual information by adopting ML algorithms (Petrović & Jovanović, 2020).

However, in the scenario of the newest generation of AI chatbots yet to be developed, it is suggested that more multimodal components be added. In the latest developments in GPT-4, the users could send visuals and textual responses. For this purpose, the use of Open AI's GPT models, Bard by Google and Blenderbot by Meta, would allow access to more advanced levels of AI chatbots operated on large language models. These would allow access to the use of multimodal chatbots in the future (Jeon, Lee & Choe, 2023; Jeon, Lee & Choi, 2023; Kohnke et al., 2023).

All in all, during the creation of AI chatbots targeting language learners specifically, application designers had better consider the characteristics of reciprocity, evaluation, personalization, omnipresence, socio-cultural knowledge, and teachableness in this conceptual model. A domain-oriented AI chatbot created for the use of language learners equipped with these qualities could offer scaffolding and explicit feedback on aspects such as vocabulary, grammar, pronunciation, accuracy, and fluency in the language with the presence of omnipresent AI avatars capable of a reciprocal interaction owing to having been created by the latest technologies available on the Internet of Things. All the same, it could also compensate for the lack of native or more proficient conversational partners. Hence, it is a crucial area for attention and further research.

5.7.4. Suggestion for an Integrated Module on AI Chatbot-assisted English Language Learning

As explained in the previous sections above, finding a freely available AI chatbot best suited to the needs of your language learners would be challenging. A pedagogically-oriented AI chatbot that allows language learners to practice language skills through a voice-enabled medium through a dialogic mode of speech is also not currently offered on the technological market. Even if there are AI chatbots created for a general audience, these applications do not initiate a task-based conversation based on thematic units, through which language learners are exposed to language units such as grammar knowledge and new vocabulary. Through goal-oriented chatbots, language learners could practice their reading, writing, listening, and speaking skills in a more guided manner in the presence of a pedagogical AI chatbot.

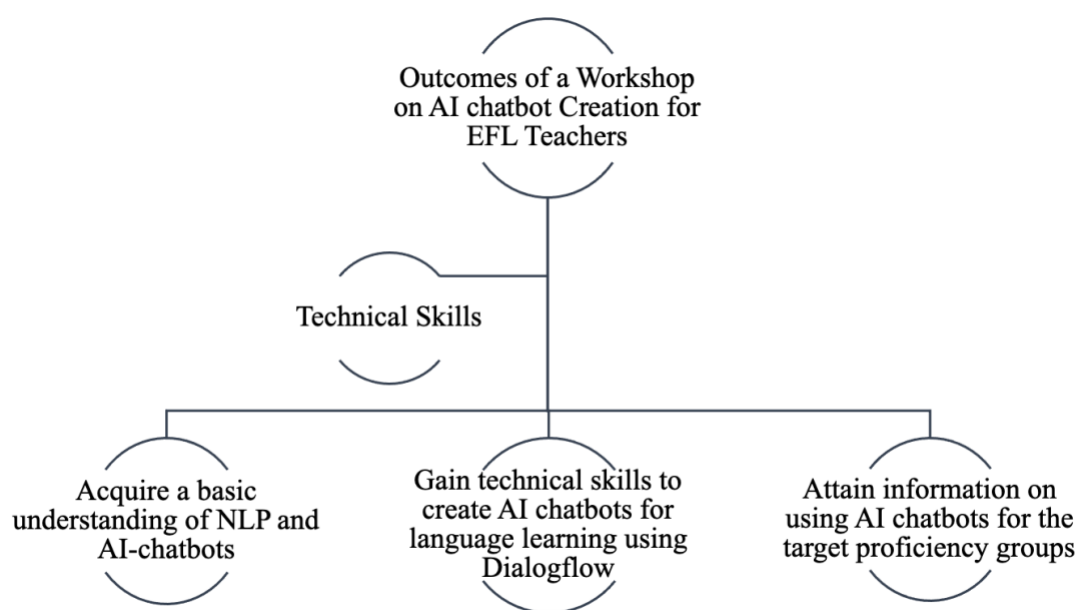


Figure 8 Objectives for a workshop on AI chatbot creation for EFL teachers

One initial step to fill this gap could be training foreign language teachers to develop AI chatbots using publicly accessible application interfaces. These interfaces, such as Google's Dialogflow, do not require any background in technical coding skills. To prepare teachers to create pedagogical AI chatbots for language learning, a workshop

could be organized through a collaboration among computer engineers, application designers, teacher educators, teachers, and language learners. As outlined in Figure 8, the intended outcomes of this workshop on AI-based chatbot development targeting language teachers would be to equip them with technical skills by first understanding the fundamental knowledge of AI-based chatbots. Teachers also would be able to learn how to use platforms such as Dialogflow that allow the creation of goal-oriented AI chatbots. They will also be supplied with the necessary information on using these tools with their specific group of learners with varying proficiency levels, needs, cultural backgrounds, and interests.

The contents of a suggested module on AI-based chatbot creation for language teachers are demonstrated in Table 56 below. It has sections presenting information about technical knowledge about AI chatbots, a dialog flow management application interface system, Dialogflow, and the pedagogical uses of these applications in the field of language teaching, as well as a practical hands-on session on the use of developed chatbots with consideration of their specific group of language learners.

Table 56 Suggested module for training on AI-based chatbot creation for EFL teachers

Part 1: Presentation		
• Presenting AI-based chatbots • The characteristics of Dialogflow • Demonstration of illustrative cases	<u>Topics:</u> • Needs analysis of language learners • Creation of pedagogical AI-based chatbots for EFL learners • Dialogue-based textual corpuses as training data for chatbots • Designing dialog flow management through the user intent, entity, response.	<u>Aims:</u> • Outline of technical knowledge on AI-powered chatbots • Attainment of information on concepts of training data, fallback, intent, entity, response

The first component in the presentation stage is introducing the characteristics of AI chatbots aimed at foreign language learners. In our case, considering that teachers working at different educational levels, from kindergarten to college, would attend this training, the characteristics of AI chatbots that would be best suitable for each language proficiency level would be discussed in this initial phase. Next, teachers would be introduced to the Dialogflow conversational flow management elements, such as interface, entity, intent, fallback, and response. These concepts are related to the training of the chatbot through entering target language structures in specific conversation scenarios, i.e., buying a train ticket at a railway station, etc., through pre-set utterances such as, “Can I buy a one-way train ticket to Piccadilly station please?”

The AI chatbot is also capable of more than talking on these subjects. The AI chatbot could learn more contextual information about the units in a language class by familiarizing the system with pre-set information. By interacting with learners, they would be capable of learning more and getting more innovative, as they are based on machine learning algorithms. What is crucial at this stage is teaching teachers how to train their AI-based chatbots based on a set of training data. In order to train chatbots, the teachers would need a textual corpus based on each intended topic, and they would manually enter this data. It is thus essential to collect an academic corpus based on English language teaching units.

Once the teachers are presented with technical information in the presentation part of the module, there will also be an implementation stage where teachers can practice what they have learned in the module (see Table 57). In this part, teachers could implement what they learned in part 1. Teachers could work individually or in groups to create an AI chatbot targeting language learners with specific proficiency levels. After creating a prototype AI chatbot, teachers would be instructed to link this chatbot to voice-enabled mobile devices such as mobile phones to activate the automatic speech recognition system and create chatbots capable of oral communication with language learners. Initially, Dialogflow allowed the integration of text-to-text output; however, it is also compatible with mobile devices to activate text-to-speech and speech-to-text qualities.

Table 57 Suggested module for training on AI-based chatbot creation for EFL teachers

Part 2: Implementation stage		
<u>Task 1:</u>	<u>Topics:</u>	<u>Aims:</u>
• Building a chatbot on Dialogflow • Integrating chatbots to voice-enabled MALL devices • Experimenting with the chatbot prototype • Demonstration of the created chatbot • Having colleague feedback	• Showing the stages of chatbot creation on the Dialogflow API • Design of conversational flow • No background on knowledge of coding is required.	• Active learning through working on training chatbots for language learners • Testing the effectiveness of newly designed chatbots on language learners
<u>Task 2:</u>		
• Testing AI-based chatbots with language learners		

After teachers develop their unique chatbots, these prototype chatbots will be tested with their language learners. Teachers, teacher educators, and application designers would take reflective notes on the experiences of language learners for further discussion at the end of the implementation session. This suggested module is argued to contribute to foreign language education by enhancing the technological pedagogical knowledge of language teachers and creating AI chatbots that are well-matched with their language teaching objectives.

5.7.5. Implications for EFL Classes and Language Teachers

Language teachers planning on utilizing AI chatbots in their language classes need to consider several factors, and there are specific steps to pursue to utilize these applications in the most rewarding way in their language classes. In incorporating AI chatbots in foreign language classes, the planning, use, and evaluation stages could be followed (see Figure 9). The first aspect of integrating AI-based chatbots in language classes is related to the planning phase, which consists of preparing the technological

infrastructures for using AI chatbots and preparing an implementation plan. AI technology integrated with AR applications requires devices with the most recent updates on mobile devices. Therefore, the teacher must ensure each student has a mobile device equipped with these features. After constructing the optimal technological settings and confirming the technological readiness of language learners, teachers must decide if the AI chatbot would be used as an in-class activity, an informal out-of-class assignment, or both.

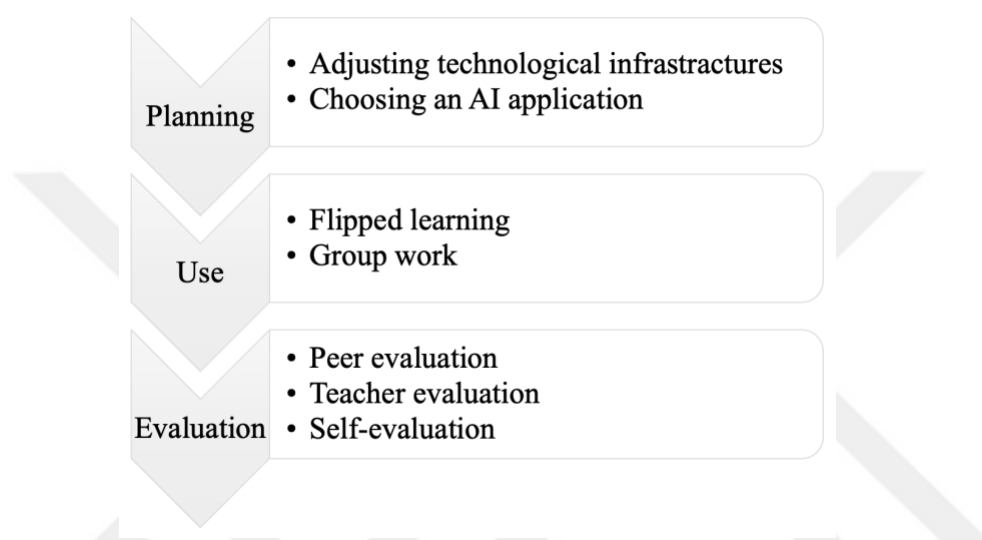


Figure 9 Phases of integration of AI-based chatbots in language classes

One suggestion for language teachers is to adopt a flipped learning strategy to integrate AI chatbots in their language classes. For instance, teachers may ask students to interact with the AI chatbot as an out-of-class activity and record their speech with the AI avatar for an in-class evaluation of the speech output through teacher or peer evaluation. Following this aim, the following activities could be flipped learning exercises.

Scenario 1: Adopting a flipped learning design, language teachers may ask learners to chat with an AI chatbot as homework and record their conversations. Students would be asked to bring the recordings to the next class. The students are put in groups of 4-5 and listen to the recorded conversations. Each student listens to a dialogue between another student and the AI chatbot. Based on a guideline their teachers delivered, learners could examine the interaction patterns of their peers on criteria such as

accuracy, fluency, vocabulary use, and spontaneity in speech. They note down the features of strengths and weaknesses and participate in a subsequent class discussion through illustrative examples with the recordings of their groups. Language learners could also get involved in a collaborative learning group activity while getting feedback from their peers on the aspects of accuracy and fluency in the language.

Scenario 2: The teacher collects the recordings of AI chatbot-language learner interactions to provide feedback to students through listening to recordings. The teacher could use these recordings as a reference point for their future teaching, as they would be able to be aware of individual students' commonly made mistakes in speaking. In a traditional classroom, giving feedback to each student might not be possible due to the class size. While listening to the interaction recordings, the teacher could note the mistakes and present the correct forms of speech in the particular areas of difficulty for this group of learners. This activity could allow language teachers to focus on improving the learners' speaking skills. Consequently, considering the findings of this study, teachers are suggested to incorporate the use of AI chatbots in both in-class and out-of-class informal learning to supplement insufficient language practice opportunities, especially in EFL settings. Thus, language teachers could make use of AI chatbots in their instruction and would make up for the time that is not always available for giving feedback to individual language learners through the support and scaffolding provided by these AI tools.

One further implication that needs attention here is the addition of a self-assessment component by language learners. In order to carry out flipped tasks efficiently, language learners should be trained on how to use AI chatbots for language learning. Researchers also suggest using a CALL training model (Dizon et al., 2022; Hubbard & Romeo, 2012) through which learners would be taught to adapt to the use of bots for language learning. This model would have technical, strategic, and pedagogical training stages. In the technical training stage, learners would be trained to use AI chatbots for language learning. Next, learners are instructed on the strategies for dealing with breakdowns in communication in the strategic training stage and informed about the reasons behind these strategies in the pedagogical training sessions. Therefore, in a flipped learning design, learners could also evaluate their language

output based on their training on the technical, strategic, and pedagogical aspects of using AI chatbots for language skills practice. If language learners could be introduced to the use of technical skills required for adopting AI tools as well as communicative strategies to carry out an effective mode of conversation, the learning benefits on their behalf could increase, too. Therefore, instructing language learners on how to best utilize AI chatbots pedagogically for each language skill needs further attention from specialists in the field. Thus, incorporating a self-assessment element into the design of teaching modules for EFL classes to evaluate language outcomes produced in interacting with AI agents is a gap in the field.

5.7.6. Implications for Application Developers

The first implication for the creators of AI chatbots for language learning and practice is a need to create pedagogical AI chatbots capable of having reciprocal interaction with students. Many publicly available chatbots for language teachers cannot usually initiate interaction while interacting with students. Consequently, students lead the conversation by initiating, taking frequent turns during AI chatbot interactions, and asking further questions. However, as the findings of this study illustrated, language learners might feel more interested in using these chatbot applications if there are more mutual questioning sequences. If not, the interaction is one way students perform monologue speeches. Moreover, the ELLs in this study interacted with a single AI avatar through a one-to-one chat. A number of language learners recounted the incorporation of the group chat feature into AI chatbots for language learning purposes. For this purpose, the application designers could create multiple avatars and enable the connection of multiple users of the AI chatbot application to connect with other avatars in an interactive open-access platform to increase the opportunities for dialogical interactions.

Another area of attention for application developers is developing an evaluation component in AI chatbot applications targeting language learners. Based on the views of the participants in the qualitative interviews and reflective diaries, and AI chatbot-student interaction data, it could be put forward that an effective AI chatbot had better have the role of an evaluator of students' mistakes through explicit error correction by

giving feedback on learner output such as pronunciation, intonation, word choice, and the use of grammar. Another suggestion for the application designers is to incorporate screen-based feedback into AI chatbots. When the speech transcription appears on screen, learners might have a chance to observe if the speech recognition ability of the system is functioning well or not and specify the reasons for any communication breakdowns (Y. Wu et al., 2020). The transcript of the dialogues is efficient because it provides immediate feedback to language learners, as they can see if the chatbot understands their speech and corrects any misunderstood words (Divekar et al., 2018). Learners might be already provided with opportunities for implicit error correction through characteristics such as screen-based feedback or transcripts of chatbots' replies. The provision of explicit feedback is one conditional feature that must necessarily be added to these types of applications.

Personalization of language learners' experience in using AI chatbots is another point that needs further development. The algorithm of the AI chatbots should be able to follow user progress and provide a personalized experience according to the student's language proficiency levels. For instance, the use of language structures by the AI chatbot needs to differentiate when interacting with a beginner or an advanced-level student. After assessing the current proficiency level of its users based on characteristics such as the words used by language learners, mean sentence length, and syntactic maturity of the sentences, the AI chatbot could assign learners to levels of language proficiency based on the CEFR levels. Adding learning materials on vocabulary, grammar, and other language skills corresponding to a variety of proficiency levels is another consideration. AI-based conversational chatbots function on mobile devices such as mobile phones, laptops, or desktop computers. Language learners can practice without time or place restrictions. One further suggestion regarding omniscience could be incorporating AI chatbot technology into wearable devices such as smartwatches or glasses and embedding these agents in virtual reality scenarios within the metaverse environment.

Creating AI chatbots capable of comprehending non-native speakers' cultural backgrounds, differences, and ways of living by integrating a sociocultural knowledge base into these applications could be another avenue for work for the creators of these

AI tools. For this purpose, the AI chatbot should be trained using non-native speaker data rich in cultural encounters collected from many countries worldwide.

While designing AI chatbots for language learning, teachableness is one property that permits the designers to develop the applications with the latest technologies available. Chatbots with innovative technologies could be constructed using the latest large language models like ChatGPT4 or more developed versions. Once AI chatbots are developed using the knowledge set in these latest progressions, the ability of AI chatbots to learn from the conversations with their users will also progress. Hence, this area needs further development by application creators. Moreover, as put forward by Jeon, Lee, Choe (2023), and Kohnke et al. (2023), depending on how large language models work in language learning environments, application designers could enhance AI-based chatbots to perform better in educational contexts in line with the particular learning aims.

5.7.7. Implications for Teacher Educators

There is a need for teacher training courses on the selection and application of chatbots in language learning contexts. Further, teaching language teachers how to create chatbots for their particular group of learners with varying learning needs could be another subject for teacher training. Teachers must also support language learners in gaining practical skills using technological tools. In this regard, H. Kim, Shin, Yang, and Lee (2019) proposed that language teachers should be equipped with the necessary knowledge to select and implement AI-based chatbots that are best suited to the needs of their students. Going one step beyond this proposition, it is suggested in this study to train language teachers to develop AI-based conversational chatbots that are best suited to their teaching environments. Thereby, a module for language teacher training on AI chatbot creation is suggested in this study for teacher development. Therefore, it is suggested that teacher educators collaborate with language teachers towards this aim.

In the chatbot development process, teachers could utilize dialog management APIs like Dialogflow, which require no background in coding skills. Alternatively, as

suggested by Jeon (2022b), chatbot developers could create basic coding models for language learning aims. Teachers could develop these models by creating chatbots with intricate communication abilities. Moreover, teachers could share the chatbots they created with their colleagues, which would further add to developing the characteristics of custom-built chatbots. However, a close collaboration with application developers, teachers, teacher educators, and language education specialists is recommended to carry out such a training module.



CHAPTER VI

CONCLUSION

6.0. Introduction

This last chapter presents an overview of this study. It provides implications for EFL classes, language teachers, application developers, and teacher educators. Limitations and suggestions for further research are discussed.

6.1. Summary of the Study

This study incorporated the results from the quantitative data and qualitative interviews, reflective diaries, and AI chatbot-language learner interactions. Implemented through a mixed-methods design, this study utilized pre-tests and post-tests to determine the listening and speaking skills achievement levels of Turkish EFL learners before and after the treatment periods. The results suggest that using AI chatbots might not yield significant differences between groups concerning listening comprehension scores and speaking proficiency development in a relatively short period. Besides its quantitative findings, this study has demonstrated various qualitative findings. The findings from the qualitative interview and reflective diary data sources have shown various opportunities and constraints of using AI chatbots for learning languages. This study provides an insight into the technological, pedagogical, and social affordances of using these tools for learning languages.

The findings of this study suggest that conversing with an AI-based chatbot in the target language allows language learners to hear target language forms with accurate pronunciation, intonation, grammar, and word choice. Vocabulary gain, boosting confidence in speakers, and enhanced grammar knowledge are some self-reported strengths of using AI chatbots for language learning. Thus, talking to an AI chatbot could enhance learners' language skills through exposure to target language forms.

Moreover, the AI chatbot-language learner interaction data in this study provide further evidence for the interaction opportunities for language learners. The findings suggest that AI chatbot use allows language learners to use various interactional strategies, i.e., semantic expansion, clarification request, and phonological modification, during a conversation to repair communication in cases of dialogical interaction. Compared to a face-to-face traditional classroom, language learners could have various opportunities to initiate conversation through AI chatbots acting as language partners.

Consequently, this study reports the properties of ideal AI chatbots for EFL learners. Based on the findings, it suggests a conceptual model, “REPOST,” consisting of the practical features for the next-generation AI chatbots. Moreover, it proposes a framework for English language teacher education for AI chatbot creation. It discusses the benefits and challenges of using AI-based chatbots for EFL learners.

6.2. Conclusion

This study is believed to contribute to the field by employing a rich array of quantitative, qualitative, and discourse analytic data sources. Each data source has yielded various outcomes pointing towards the affordances of AI chatbots for learning languages. First of all, the quantitative findings in this study showed insignificant differences in terms of listening and speaking skills development between the experimental and the control group. However, these insignificant results could be the result of several factors. First, the insignificant differences could be due to the implementation period of the study. This study was conducted for two months, which could have led to insignificant differences in the language achievement test scores. Furthermore, a novelty impact could have yielded insignificant results in the two groups' listening and speaking proficiency test scores. At the beginning of the implication period, learners might have demonstrated an interest in this novel technological learning tool; however, this initial interest in using the AI chatbot application might have dropped after a particular duration of use.

Concerning the pedagogical opportunities that were brought about through AI chatbots in language education, through the use of an AI chatbot, language learners in this EFL context could have a speaking partner to practice language skills in the target language. Thus, they could be exposed to the target language patterns via the voice-call feature embedded in this AI chatbot application, which the previous generation of AI chatbots lacked, as the previous chatbot applications were primarily text-based, allowing a textual form of connection with its users. This study is an endeavor to fill the gap in the literature regarding the need for studies on the implementations of voice-enabled AI chatbots in EFL learners' speaking skills. The literature on writing skills development with AI generative technologies has surged considerably compared to studies exploring the benefits of voice-based AI technologies (Liu et al., 2021; Yang et al., 2023; Zhao, 2022). Therefore, the results of this study would form the ground for more empirical studies to be conducted on the use of AI for speaking skills practice in foreign languages.

An area that needs attention here is that although learners in this study stated that there was limited feedback from the AI chatbot, they could be implicitly exposed to error correction. Namely, although the AI chatbot did not explicitly give feedback to students on their ungrammatical sentences or mispronounced words, students utilized many strategies when there was a communication breakdown, considering that their use of target language structures led to a misunderstanding. Consequently, the production of modified output led to an enhanced awareness of the target language forms through implicit error correction by the language learners.

The affordances also impacted the psychological characteristics of students either positively or negatively depending on the nature of their interaction experiences in this study. The psychological sides specified in this study were self-reports by learners regarding L2 confidence, elevated motivation, and decreased speaking anxiety. The social affordances revealed in this study have implications for the implementations of AI chatbots in alleviating language learning anxiety of learners as an impact of chatting with a nonhumane AI avatar.

The language learners' beliefs in this study pointed towards the benefits of AI chatbots in increasing their confidence to speak in the target language. As learners were aware that they were interacting with a conversation partner free of judgment, they could feel relaxed enough to express themselves more freely in the foreign language without an evaluative listener during the interactions. These components supplied implications concerning how AI chatbots could be integrated into language classes or employed as independent learning tools.

6.3. Limitations and Suggestions for Further Research

The quantitative results of this study have demonstrated that no results were pointing to significant differences in the use of AI chatbots on learning outcomes between the experimental and the control groups in this study. This could be because this study was implemented within two months in an academic semester. Studies conducted collecting longitudinal data would provide a broader spectrum of the representation of the learning outcomes of English language learners.

Another limitation of this study was related to the mode of education the educational institution was implementing at the time of the study. A hybrid education model was in place as a part of the COVID-19 disease prevention measurements taken by the rectorate of the higher education institution in which this study was conducted. Therefore, many of the students did not participate in face-to-face lectures. Some of these students needed easy access to technological facilities such as a mobile device that was the most up-to-date or a strong internet connection, which led to a relatively low participation rate among the students.

However, according to the duration of the recorded interaction data and student self-reports, they used the application for more than approximately 20 minutes each week and up to one hour in some instances in order to come up with a coherent conversation to be uploaded into our educational management system for the access of the researcher. Therefore, further studies might be carried out in other educational contexts equipped with more high-tech devices that would have even higher participation rates among the participants.

Suggestions for further research concern the investigation of the negative impacts and evaluations on the adoption of AI-based conversational agents, as there are few accounts of negative experiences of language learners in the literature. Many studies report the positive outcomes of using chatbot technology for practicing second or foreign language skills. However, as reported in this study, AI chatbot use has mixed impacts on learning after-effects. Moreover, other variables, such as the impacts of learning styles and cultural differences on the learning outcomes after AI chatbot implementations, could be explored in future studies.

Future researchers may also investigate whether the chatbots developed through application interfaces (API) connected to the most recent large language models affect the language learning process. It would be an essential area of study exploring how teachers would create chatbots using entirely functioning dialogue management systems with the components of entities, prompts, and intents that would promote learner interaction and use of metacognitive strategies (Jeon, Lee & Choe, 2023; Kohnke et al., 2023).

Another area for the study could be exploring the effects of chatbots adopting different modalities, such as text-based, voice-based, and multimodal aspects, on learning outcomes. The data could be collected through different means, such as users' self-ratings on the application's effectiveness, language learning outcomes in various domains, and interactional data. Thus, future studies could explore whether multimodalities impact language learning.

Jeon, Lee, and Choe (2023) also proposed that as graphical design technology progresses, how close a chatbot resembles a human being will be an essential element that could impact learner-chatbot conversation and its impact on learning. Thus, the effect of incorporating human-like features like personality and bodily movements such as gestures on large language models is yet to be discovered. Interaction with 'virtually embedded chatbots' would become more possible in the future with the developments in metaverse technology, and these areas would provide significant avenues for investigation with a focus on the interaction between virtually embedded AI chatbots and their users in the metaverse.

REFERENCES

- Ahn, S. (2022). The effects of chatbot on grammar competence for Korean EFL college students. *Journal of Digital Convergence*, 20(3), 53–61. <https://doi.org/10.14400/JDC.2022.20.3.053>
- Ali Amer Jid Almahri, F., Bell, D., Arzoky, M, 2019. Personas design for conversational systems in education. *Informatics* 6 (4), 46. <https://doi.org/10.3390/informatics6040046>
- Amaral, L. A., & Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, 23(1), 4–24. <https://doi.org/10.1017/S0958344010000261>
- Argal, A., Gupta, S., Modi, A., Pandey, P., Shim, S., & Choo, C. (2018, January). Intelligent travel chatbot for predictive recommendation in echo platform. In *2018 IEEE 8th annual computing and communication workshop and conference (CCWC)* (pp. 176-183). IEEE.
- Ayedoun, E., Hayashi, Y., & Seta, K. (2019). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *International Journal of Artificial Intelligence in Education*, 29, 29-57. <https://doi.org/10.1007/s40593-018-0171-6>
- Ayedoun, E., Hayashi, Y., & Seta, K. (2020). Toward personalized scaffolding and fading of motivational support in L2 learner-dialogue agent interactions: An exploratory study. *IEEE Transactions on Learning Technologies*, 13(3), 604–616. <https://doi.org/10.1109/TLT.2020.2989776>
- Bachman, L. (1990) *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bailey, D. (2019). Chatbots as conversational agents in the context of language learning. In *Proceedings of the Fourth Industrial Revolution and Education* (pp. 32–41).
- Bailey, D., Southam, A., & Costley, J. (2021). Digital storytelling with chatbots: Mapping L2 participation and perception patterns. *Interactive Technology and Smart Education*, 18(1), 85–103. <https://doi.org/10.1108/ITSE-08-2020-0170>
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, 12(17), 8427. <https://doi.org/10.3390/app12178427>

- Best, J. W., Kahn, J. V., & Jha, A. K. (2017). *Research in education*. Pearson Education Limited.
- Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning*, 32(8), 827–877. <https://doi.org/10.1080/09588221.2018.1535508>
- Bibauw, S., François, T., & Desmet, P. (2022). Dialogue systems for language learning: Chatbots and beyond. *The Routledge handbook of second language acquisition and technology*, 121-134. <https://doi.org/10.4324/9781351117586-12>
- Bibauw, S., François, T., Van den Noortgate, W., & Desmet, P. (2022). Dialogue systems for language learning: A meta-analysis. *Language, Learning and Technology*, 26 (1), 1–24. <https://hdl.handle.net/10125/73488>
- Bower, J., & Kawaguchi, S. (2011). Negotiation of meaning and corrective feedback in Japanese/English eTandem. *Language Learning & Technology*, 15(1), 41–71. <https://www.lltjournal.org/item/2723>.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1–47. <https://doi.org/10.1093/applin/I.1.1>
- Cancino, M. (2015). Assessing learning opportunities in EFL classroom interaction: What can conversation analysis tell us?. *RELC Journal*, 46(2), 115-129. <https://doi.org/10.1177/0033688214568109>
- Chang, C. Y., Kuo, S. Y., & Hwang, G. H. (2022). Chatbot-facilitated nursing education: Incorporating a knowledge-based chatbot system into a nursing training program. *Educational Technology and Society*, 25(1), 15-27.
- Chapelle, C. (2001). *Computer applications in second language acquisition : Foundations for teaching, testing and research*. Cambridge University Press.
- Chapelle, C. A. (2006). Interactionist SLA theory in CALL research. In J. L. Egbert & G. M. Petrie (Eds.), *CALL Research Perspectives* (pp.53–64). Mahwah, NJ: Lawrence Erlbaum.
- Chen, J. C. (2018). The interplay of tasks, strategies and negotiations in Second Life. *Computer Assisted Language Learning*, 31(8), 960-986. <https://doi.org/10.1080/09588221.2018.1466810>
- Chen, H. L., Vicki Widarso, G., & Sutrisno, H. (2020). A chatbot for learning Chinese: Learning achievement and technology acceptance. *Journal of Educational Computing Research*, 58(6), 1161–1189. <https://doi.org/10.1177/0735633120929622>

- Chen, H. H. J., Yang, C. T. Y., & Lai, K. K. W. (2023). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments*, 31(3), 1335-1350. <https://doi.org/10.1080/10494820.2020.1833043>
- Chien, Y. C., Wu, T. T., Lai, C. H., & Huang, Y. M. (2022). Investigation of the influence of artificial intelligence markup language-based LINE ChatBot in contextual English Learning. *Frontiers in Psychology*, 13. Article 785752. <https://doi.org/10.3389/fpsyg.2022.785752>
- Choi, S. K. Kwon, O. W. & Kim, Y. K. (2017). Computer-assisted English learning system based on free conversation by topic. In K. Brothwick, L. Bradley & Thouseny, *CALL in a climate of change: adapting to turbulent global conditions. Short papers from EUROCALL 2017*, 79-85.
- Chuah, K. M., & Kabilan, M. K. (2021). Teachers' views on the use of chatbots to support English language teaching in a mobile environment. *International Journal of Emerging Technologies in Learning*, 16(20), 223–237. <https://doi.org/10.3991/ijet.v16i20.24917>
- Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, 117, 587-595. <https://doi.org/10.1016/j.jbusres.2018.10.004>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education*. Routledge.
- Coniam, D. (2008a). An evaluation of chatbots as software aids to learning English as a second language. *The EUROCALL Review*, 13, 2–14. <https://doi.org/10.4995/eurocall.2008.16353>
- Coniam, D. (2008b). Evaluating the language resources of chatbots for their potential in English as a second language. *ReCALL*, 20(1), 98–116. <https://doi.org/10.1017/S0958344008000815>
- Coniam, D. (2014). The linguistic accuracy of chatbots: Usability from an ESL perspective. *Text and Talk*, 34(5), 545–567. <https://doi.org/10.1515/text-2014-0018>
- Council of Europe, Council for Cultural Co-operation, Education Committee, Modern Languages Division. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage publications.
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research*. Sage publications.
- Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017, July). Superagent: A customer service chatbot for e-commerce websites. In *Proceedings of ACL 2017, System Demonstrations* (pp. 97-102).
- Cutrone, P. (2009). Overcoming Japanese EFL learners' fear of speaking. *University of Reading Language Studies Working Papers*, 1(1), 55-63. Retrieved from http://www.reading.ac.uk/AcaDepts/ll/app_ling/internal/Cutrone.pdf246 P.L. Ha and B. Li
- Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, 22(5), 811-817. <https://doi.org/10.1017/S1351324916000243>
- Danli, L. (2017). Autonomy in scaffolding as learning in teacher-student negotiation of meaning in a university EFL classroom. *Chinese Journal of Applied Linguistics*, 40(4), 410-430. <https://doi.org/10.1515/cjal-2017-0024>
- Daniels, P., & Iwago, K. (2017). The suitability of cloud-based speech recognition engines for language learning. *JALT CALL Journal*, 13(3), 229–239.
- Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods*. Prentice Hall.
- Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., Zhu, H., Hendler, J. A., & Braasch, J. (2021). Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation. *Computer Assisted Language Learning*. 35(9), 2332-2360. <https://doi.org/10.1080/09588221.2021.1879162>
- Divekar, R. R., Drozdal, J., Zhou, Y., Song, Z., Allen, D., Rouhani, R., Zhao, R., Zheng, S., Balagyozyan, L., & Su, H. (2018). Interaction challenges in AI equipped environments built to teach foreign languages through dialogue and task-completion. *DIS 2018 - Proceedings of the 2018 Designing Interactive Systems Conference*, 597–610. <https://doi.org/10.1145/3196709.3196717>
- Divya, S., Indumathi, V., Ishwarya, S., Priyasankari, M., & Devi, S. K. (2018). A self-diagnosis medical chatbot using artificial intelligence. *Journal of Web Development and Web Designing*, 3(1), 1-7.
- Diwanji, P., Hinkelmann, K., & Witschel, H. (2018). Enhance classroom preparation for flipped classroom using AI and analytics. In S. Hammoudi, M. Smialek, O. Camp, & J. Filipe (Eds.), *Proceedings of the 20th international conference on*

- enterprise information systems (ICEIS 2018)* (Vol. 1, pp. 477–483). Madeira, Portugal: SciTePress. <https://doi.org/10.5220/0006807604770483>
- Dizon, G. (2017). Using intelligent personal assistants for second language learning: A case study of Alexa. *TESOL Journal*, 8(4), 811–830. <https://doi.org/10.1002/tesj.353>
- Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language, Learning and Technology*, 24(1), 16–26.
- Dizon, G. (2021). Affordances and constraints of intelligent personal assistants for second-language learning. *RELC Journal*. Article 003368822110205. <https://doi.org/10.1177/00336882211020548>,
- Dizon, G., & Tang, D. (2020). Intelligent personal assistants for autonomous second language learning: An investigation of Alexa. *JALT CALL Journal*, 16(2), 107–120. <https://doi.org/10.29140/jaltcall.v16n2.273>
- Dizon, G., Tang, D., & Yamamoto, Y. (2022). A case study of using Alexa for out-of-class, self-directed Japanese language learning. *Computers and Education: Artificial Intelligence*, 3. <https://doi.org/10.1016/j.caeai.2022.100088>
- Dodigovic, M. (2005). *Artificial intelligence in second language learning*. Multilingual Matters.
- Dokukina, I., & Gumanova, J. (2020). The rise of chatbots – New personal assistants in foreign language learning. *Procedia Computer Science*, 169, 542–546. <https://doi.org/10.1016/j.procs.2020.02.212>
- Donald, S. (2020). Learner initiatives in the EFL classroom: A public/private phenomenon. *ELT Journal*, 74(2), 136–145. <https://doi.org/10.1093/elt/ccaa006>
- Doughty, C. (2000). Negotiating the L2 linguistic environment. *University of Hawaii Working Papers in ESL*, 18, 47–83.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum Associates.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- Durall, E., & Kapros, E. (2020, July). Co-design for a competency self-assessment chatbot and survey in science education. In *International Conference on Human-Computer Interaction* (pp. 13–24). Springer, Cham.
- Ebadi, S., & Amini, A. (2022). Examining the roles of social presence and human-likeness on Iranian EFL learners' motivation using artificial intelligence technology: A case of CSIEC chatbot. Advance online publication. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2022.2096638>

- Elaish, M. M., Shuib, L., Ghani, N. A., & Yadegaridehkordi, E. (2017). Mobile English language learning (MELL): A literature review. *Educational Review*, 71(2), 257-276. <https://doi.org/10.1080/00131911.2017.1382445>
- El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, Article e12667. 38(3). <https://doi.org/10.1111/exsy.12667>
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Ericsson, E., Sofkova Hashemi, S., & Lundin, J. (2023). Fun and frustrating: Students' perspectives on practising speaking English with virtual humans. *Cogent Education*, 10(1), 2170088. <https://doi.org/10.1080/2331186X.2023.2170088>
- Fernández-García, M., & Martínez-Arbelaiz, A. (2002). Negotiation of meaning in nonnative speaker-nonnative speaker synchronous discussions. *CALICO Journal*, 19(2), 279-294. <https://www.jstor.org/stable/24149362>
- Foster, P. (1998). A classroom perspective on negotiation for meaning. *Applied Linguistics*, 19(1), 1–23. <https://doi.org/10.1093/applin/19.1.1>
- Foster, P., & Ohta, A. (2005). Negotiation for meaning and peer assistance in second language classrooms. *Applied Linguistics*, 26(3), 402–430. <https://doi.org/10.1093/applin/ami014>
- Fryer, L. K. (2019). Getting interested: Developing a sustainable source of motivation to learn a new language at school. *System*, 86, Article 102120 <https://doi.org/10.1016/j.system.2019.102120>
- Fryer, L.K., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology*, 10(3), 8-14.
- Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior*, 75, 461–468. <https://doi.org/10.1016/j.chb.2017.05.045>
- Fryer, L.K., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology*, 24 (2), 8-22. Retrieved from <http://hdl.handle.net/10125/44719>
- Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*, 93, 279–289. <https://doi.org/10.1016/j.chb.2018.12.023>
- Fryer, L. K., Thompson, A., Nakao, K., Howarth, M., & Gallacher, A. (2020). Supporting self-efficacy beliefs and interest as educational inputs and outcomes: Framing AI and human partnered task experiences. *Learning and Individual Differences*, 80, Article 101850. <https://doi.org/10.1016/j.lindif.2020.101850>

- Garcia-Ponce, E. E., & Mora-Pablo, I., (2018). Incidence and nature of negotiations for meaning during uncontrolled speaking practice in English as a foreign language classrooms. *Journal of Language and Education*, 4(1), 31-41. <https://doi.org/10.17323/2411-7390-2018-4-1-31-41>
- Gass, S., & Selinker, L. (1994). *Second language acquisition: An introductory course*. Lawrence Erlbaum Associates.
- George, D., & Mallery, P. (2016). *IBM SPSS Statistics 23 step by step: A simple guide and reference* (14th ed.). Routledge.
- Ghorbani, N., & Ebadi, S. (2020). Exploring learners' grammatical development in mobile assisted language learning. *Cogent Education*, 7(1), Article 1704599.
- Given, L.M. & Saumure, K. (2008). Trustworthiness. In L. M. Given (Ed.), *The Sage encyclopedia of qualitative research methods* (pp. 139-140). Thousand Oaks, Sage Publishing. <http://dx.doi.org/10.4135/9781412963909.n77>
- Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social problems*, 12(4), 436-445.
- Goda, Y., Yamada, M., Matsukawa, H., Hata, K., & Yasunami, S. (2014). Conversation with a chatbot before an online EFL group discussion and the effects on critical thinking. *The Journal of Information and Systems in Education*, 13(1), 1-7
- Gonulal, T. (2021). Investigating EFL learners' humorous interactions with an intelligent personal assistant. *Interactive Learning Environments*, 31(7), 1-14. <https://doi.org/10.1080/10494820.2021.1974489>
- Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing*, 54, 1-6. Article 100666. <https://doi.org/10.1016/j.asw.2022.100666>
- Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Investigating students' engagement in chatbot-supported classroom debates. *Interactive Learning Environments*. 1-17. <https://doi.org/10.1080/10494820.2023.2207181>
- Han, D.E. (2020). The effects of voice-based AI chatbots on Korean EFL middle school students' speaking competence and affective domains. *Asia-Pacific Journal of Convergent Research Interchange*, 6(7), 71-80. <https://doi.org/10.47116/apjcri.2020.07.07>
- Hapsari, I. P., Wu, T. T. (2022). AI chatbots learning model in English speaking skill: Alleviating speaking anxiety, boosting enjoyment, and fostering critical thinking. *International Conference on Innovative Technologies and Learning* (pp. 444-453). Springer.

- Haristiani, N. (2019). Artificial intelligence (AI) chatbot as language learning medium: An inquiry. *Journal of Physics: Conference Series*, 1387(1). 1-6. <https://doi.org/10.1088/1742-6596/1387/1/012020>
- Haristiani, N., Dewanty, V. L., & Rifai, M. M. (2022). Autonomous learning through chatbot-based application utilization to enhance basic Japanese competence of vocational high school students. *Journal of Technical Education and Training*, 14(2), 143–155. <https://doi.org/10.30880/jtet.2022.14.02.013>
- Heller, B., Proctor, M., Mah, D., Jewell, L., & Cheung, B. (2005, June). Freudbot: An investigation of chatbot technology in distance education. In *EdMedia+ Innovate Learning* (pp. 3913-3918). Association for the Advancement of Computing in Education (AACE).
- Hew, K. F., Huang, W., Du, J., & Jia, C. (2022). Using chatbots to support student goal setting and social presence in fully online activities: learner engagement and perceptions. *Journal of Computing in Higher Education*. 35, 40-68. <https://doi.org/10.1007/s12528-022-09338-x>
- Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Computers in Human Behavior*, 49, 245–250. <https://doi.org/10.1016/j.chb.2015.02.026>
- Ho, Y. K. (2003). Audiotaped dialogue journals: An alternative form of speaking practice. *ELT Journal*, 57(3), 269-277. <https://doi.org/10.1093/elt/57.3.269>
- Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, 68(4), 712-733. <https://doi.org/10.1093/joc/jqy026>
- Hong, Z.W., Huang, Y.M., Hsu, M., & Shen, W.-W. (2016). Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL classrooms. *Journal of Educational Technology & Society*, 19(1). 337-349. <https://doi.org/10.2307/jeductechsoci.19.1.337>
- Hsu, L. (2022). To CALL or not to CALL: empirical evidence from neuroscience. *Computer Assisted Language Learning*, 35(4), 792–815. <https://doi.org/10.1080/09588221.2020.1750429>
- Hsu, H. L., Chen, H. H. J., & Todd, A. G. (2023). Investigating the impact of the Amazon Alexa on the development of L2 listening and speaking skills. *Interactive Learning Environments*, 31(9), 5732-5745. <https://doi.org/10.1080/10494820.2021.2016864>
- Hsu, M. H., Chen, P. S., & Yu, C. S. (2023). Proposing a task-oriented chatbot system for EFL learners speaking practice. *Interactive Learning Environments*, 31(7), 4297-4308. <https://doi.org/10.1080/10494820.2021.1960864>

- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), 237–257. <https://doi.org/10.1111/jcal.12610>
- Hubbard, P., & Romeo, K. (2012). Diversity in learner training. In G. Stockwell (Ed.), *Computer-assisted language learning: Diversity in research and practice* (pp. 33–48). Cambridge University Press.
- Hwang, W. Y., Guo, B. C., Hoang, A., Chang, C. C., & Wu, N. T. (2022). Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating its influence on learning achievements. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2022.2095406>
- Istrate, A. M. (2019). The impact of the virtual assistant (VA) on language classes. In *Proceedings of the 15th International Scientific Conference eLearning and Software for Education* (pp. 296–301). Bucharest, Romania: Carol I National Defence University.
- Jensen, D. (2008). Credibility. In L. M. Given (Ed.), *The Sage encyclopedia of qualitative research methods* (pp. 139-140). Thousand Oaks, Sage Publishing. <http://dx.doi.org/10.4135/9781412963909.n77>
- Jeon, J. (2022a). Exploring a self-directed interactive app for informal EFL learning: a self-determination theory perspective. *Education and Information Technologies*, 27(4), 5767–5787. <https://doi.org/10.1007/s10639-021-10839-y>
- Jeon, J. (2022b). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2021.2021241>
- Jeon, J. (2023). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning*, 36(7), 1338-1364. <https://doi.org/10.1080/09588221.2021.1987272>
- Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, 206, 104898. <https://doi.org/10.1016/j.compedu.2023.104898>
- Jeon, J., Lee, S., & Choi, S. (2023). A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models. *Interactive Learning Environments*. Advance online publication. <https://doi.org/10.1080/10494820.2023.2204343>
- Jepson, K. (2005). Conversations—and negotiated interaction—in text and voice chat rooms. *Language Learning & Technology*, 9(3), 19–98.

- Jia, J. (2004). The study of the application of a web-based chatbot system on the teaching of foreign languages. In *Society for Information Technology & Teacher Education International Conference* (pp. 1201-1207). Association for the Advancement of Computing in Education (AACE).
- Jia, J. (2009). CSIEC: A computer assisted English learning chatbot based on textual knowledge and reasoning. *Knowledge-Based Systems*, 22(4), 249-255. <https://doi.org/10.1016/j.knosys.2008.09.001>
- Jia, J. & Chen, W. (2008) Motivate the learners to practice English through playing with chatbot CSIEC. In Pan, Z., Zhang, X., El Rhalibi, A., Woo, W. & Li, Y. (eds.), *Technologies for e-learning and digital entertainment: Edutainment 2008: Vol. 5093: Lecture notes in computer science*. Berlin: Springer, 180–191. https://doi.org/10.1007/978-3-540-69736-7_20
- Jia, J., & Chen, W. (2009). The further development of CSIEC project driven by application and evaluation in English education. *British Journal of Educational Technology*, 40(5), 901–918. <https://doi.org/10.1111/j.1467-8535.2008.00881.x>
- Johnson, W. L. (2019). Data-driven development and evaluation of Enskill English. *International Journal of Artificial Intelligence in Education*, 29(3), 425–457. <https://doi.org/10.1007/s40593-019-00182-2>
- Junaidi, J., Hamuddin, B., Julita, K., Rahman, F., & Derin, T. (2020). Artificial Intelligence in EFL Context: Rising Students' Speaking Performance with Lyra Virtual Assistance. *International Journal of Advanced Science and Technology*, 29(5), 6735–6741. Retrieved from <http://sersc.org/journals/index.php/IJAST/article/view/17726>
- Kang, H. (2022). Effects of artificial intelligence (AI) and native speaker interlocutors on ESL learners' speaking ability and affective aspects. *Multimedia-Assisted Language Learning*, 25(2), 9–43. <https://doi.org/10.15702/mall.2022.25.2.09>
- Kara, E., Ayaz, A. D., & Dündar, T. (2017). Challenges in EFL speaking classes in Turkish context. *European Journal of Language and Literature*, 3(2), 66-74.
- Katsarou, E., Wild, F., Sougari, A. M., & Chatzipanagiotou, P. (2023). A systematic review of voice-based intelligent virtual agents in EFL education. *International Journal of Emerging Technologies in Learning*, 18(10), 65–85. <https://doi.org/10.3991/ijet.v18i10.37723>
- Kaur, A., Singh, S., Chandan, J. S., Robbins, T., & Patel, V. (2021). Qualitative exploration of digital chatbot use in medical education: A pilot study. *Digital Health*, 7, 20552076211038151. <https://doi.org/10.1177/20552076211038151>
- Kent, D. (2021). Voice-user interfaces for TESOL: Potential and receptiveness among native and non-native English-speaking instructors. *Language Learning & Technology*, 25(3), 27-39. Taken from <http://hdl.handle.net/10125/73444>

- Kessler, G. (2018). Technology and the future of language teaching. *Foreign Language Annals*, 51(1), 205–218. <https://doi.org/10.1111/flan.12318>
- Kılıçkaya, F. (2020). Using a chatbot, Replika, to practice writing through conversations in L2 English: A Case study. In M. Kruk, & M. Peterson, (Eds.), *New Technological Applications for Foreign and Second Language Learning and Teaching* (pp. 221-238). <https://doi.org/10.4018/978-1-7998-2591-3.ch011>
- Kim, I. S. (2006). Automatic Speech Recognition: Reliability and Pedagogical Implications for Teaching Pronunciation. *International Forum of Educational Technology & Society*, 9(1), 322–334. <https://doi.org/10.2307/jeductechsoci.9.1.322>
- Kim, N. -Y. (2016). Effects of voice chat on EFL learners' speaking ability according to proficiency levels. *Multimedia-Assisted Language Learning*, 19(4), 63–88.
- Kim, N. -Y. (2017). Effects of types of voice-based chat on EFL students' negotiation of meaning according to proficiency levels. *English Teaching*, 72(1), 159–181.
- Kim, N. -Y. (2018a). A study on chatbots for developing Korean college students' English listening and reading. *Journal of Digital Convergence*, 16(8), 19–26.
- Kim, N. -Y. (2018b). Chatbots and Korean EFL students' English vocabulary learning. *Journal of Digital Convergence*, 16(2), 1–7. <https://doi.org/10.14400/JDC.2018.16.2.001>
- Kim, N. -Y. (2019). A study on the use of artificial intelligence chatbots for improving English grammar skills. *Journal of Digital Convergence*, 17(8), 37–46. <https://doi.org/10.14400/JDC.2019.17.8.037>
- Kim, H., Shin, D. K., Yang, H. & Lee, J. H. (2019) A study of AI chatbot as an assistant tool for school English curriculum. *Journal of Learner-Centered Curriculum and Instruction*, 19(1), 89–110. <https://doi.org/10.22251/jlcci.2019.19.1.89>
- Kim, H. S., Cha, Y., & Kim, N. -Y. (2021). Effects of AI chatbots on EFL students' communication skills. *Korean Journal of English Language and Linguistics*, 2021(21), 712–734. <https://doi.org/10.15738/kjell.21..202108.712>
- Kim, H. S., Kim, N. -Y., & Cha, Y. (2021). Is it beneficial to use AI chatbots to improve learners' speaking performance? *Journal of Asia TEFL*, 18(1), 161–178. <https://doi.org/10.18823/asiatefl.2021.18.1.10.161>
- Kim, N.-Y., Cha, Y., & Kim, H.-S. (2019). Future English learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning*, 22(3), 32–53.
- Kohnke, L. (2023). L2 learners' perceptions of a chatbot as a potential independent language learning tool. *Int. J. Mobile Learning and Organisation*, 17, 214–226. <https://doi.org/10.1504/IJMLO.2023.10053355>

- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537-550.
<https://doi.org/10.1177/00336882231162868>
- Krashen, S. D. (1981). *Second language acquisition and second language learning*. Pergamon Press.
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28, 973-1018. <https://doi.org/10.1007/s10639-022-11177-3>
- Kukulska-Hulme, A. (2020). Mobile-Assisted Language Learning. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1–9). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781405198431.wbeal0768.pub2>
- Kukulska-Hulme, A., & Lee, H. (2020). Intelligent assistants in language learning: an analysis of features and limitations. In *CALL for widening participation: short papers from EUROCALL 2020* (pp. 172–176). Research-publishing.net. <https://doi.org/10.14705/rpnet.2020.48.1184>
- Kuligowska, K. (2015) Commercial chatbot: Performance evaluation, usability metrics and quality standards of embodied conversational agents. *Professionals Center for Business Research*, 2(2),1-16. <https://doi.org/10.18483/PCBR.22>
- Kusuma, P. C., Indriyani, C. E., & Xenia, T. (2023). An analysis of the negotiation of meaning and form in natural EFL classrooms: A case study among pre-service teachers. *LLT Journal: A Journal on Language and Language Teaching*, 26(1), 54-68. <https://doi.org/10.24071/llt.v26i1.4799>
- Lee, L. (2007). Fostering second language oral communication through constructivist interaction in desktop videoconferencing. *Foreign Language Annals*, 40(4), 635–649. <https://doi.org/10.1111/j.1944-9720.2007.tb02885.x>
- Lee, S., & Jeon, J. (2022). Visualizing a disembodied agent: Young EFL learners' perceptions of voice-controlled conversational agents as language partners. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2022.2067182>
- Lee, S., Noh, H., Lee, J., Lee, K., Lee, G. G., Sagong, S., & Kim, M. (2011). On the effectiveness of robot-assisted language learning. *ReCALL*, 23(1), 25-58. <https://doi.org/10.1017/S0958344010000273>
- Lee, J. H., Yang, H., Shin, D., & Kim, H. (2020). Chatbots. *ELT Journal*. 74(3), 338–344.
- Leech, N. L., & Onwuegbuzie, A. J. (2007). An array of qualitative data analysis tools: A call for data analysis triangulation. *School psychology quarterly*, 22(4), 557-584.

- Lin, M. P.-C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot. *Journal of Educational Technology & Society*, 23(1), 78–92. <https://doi.org/10.2307/26915408>
- Lin, C.J., & Mubarak, H. (2021). Learning Analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society*, 24(4), 16–35. <https://doi.org/10.2307/48629242>
- Liu, X. (2022, April). The relation of input, interaction, and output in SLA. In *2022 7th International Conference on Social Sciences and Economic Development (ICSSED 2022)* (pp. 474-479). Atlantis Press.
- Liu, C., Hou, J., Tu, Y. F., Wang, Y., & Hwang, G. J. (2023). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments*, 31(9), 5614-5632. <https://doi.org/10.1080/10494820.2021.2012812>
- Liu, C. C., Liao, M. G., Chang, C. H., & Lin, H. M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers and Education*, 189. Article 104576 <https://doi.org/10.1016/j.compedu.2022>.
- Long, M. H. (1983). Linguistic and conversational adjustment to non-native speakers. *Studies in Second Language Acquisition*, 5(2), 177-193.
- Long, M. H. (1990). The least a second language acquisition theory needs to explain. *TESOL Quarterly*, 24(4), 649–666. <https://doi.org/10.2307/3587113>
- Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback?. *Studies in second language acquisition*, 22(4), 471-497. <https://doi.org/10.1017/S0272263100004022>
- Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI chatbots for content and language integrated learning. *Applied Sciences*, 12(7). Article 3239. <https://doi.org/10.3390/app12073239>
- Malik, G., Tayal, D. K., & Vij, S. (2019). An analysis of the role of artificial intelligence in education and teaching. In *Recent Findings in Intelligent Computing Techniques* (pp. 407-417). Springer, Singapore.
- McStay, A. (2023). Replika in the metaverse: The moral problem with empathy in 'It from Bit'. *AI and Ethics*, 3(4), 1433-1445. <https://doi.org/10.1007/s43681-022-00252-7>
- Meri-Yilan, S. (2019) A study on technology-based speech assistants. In A. Plutino, K. Borthwick & E. Corradini (Eds), *New educational landscapes: Innovative perspectives in language learning and technology* (pp. 11-17). Research-publishing.net

- Miles, M. B. & Huberman, A. M. (1984). *Qualitative data analysis*. Sage.
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (Second edition). Sage.
- Mohamed, A. M. (2023). Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching: perceptions of EFL faculty members. *Education and Information Technologies*. Advance online publication. <https://doi.org/10.1007/s10639-023-11917-z>
- Moradi, A., & Farvardin, M. T. (2020). Negotiation of meaning by mixed-proficiency dyads in face-to-face and synchronous computer-mediated communication. *TESOL Journal*, 11(1), <https://doi.org/10.1002/tesj.446>
- Morton, H., & Jack, M. (2010). Speech interactive computer-assisted language learning: A cross-cultural evaluation. *Computer Assisted Language Learning*, 23(4), 295-319. <https://doi.org/10.1080/09588221.2010.493524>
- Moussalli, S., & Cardoso, W. (2016). Are commercial ‘personal robots’ ready for language learning? Focus on second language speech. In S. Papadima-Sophocleous, L. Bradley, & S. Thouesny (Eds.), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 325–329). Research-publishing.net.
- Moussalli, S., & Cardoso, W. (2020). Intelligent personal assistants: Can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*, 33(8), 865-890. <https://doi.org/10.1080/09588221.2019.1595664>
- Moussalli, S., & Cardoso, W. (2021). Intelligent personal assistants and L2 pronunciation development: Focus on English past -ed. In *CALL and professionalisation: short papers from EUROCALL 2021* (pp. 226–231). Research-publishing.net. <https://doi.org/10.14705/rpnet.2021.54.1337>
- Nagata, R., Hashiguchi, T., & Sadoun, D. (2020). Is the simplest chatbot effective in English writing learning assistance? In *International Conference of the Pacific Association for Computational Linguistics* (pp. 245–256). Springer, Singapore. https://doi.org/10.1007/978-981-15-6168-9_21.
- Nghi, T. T., Phuc, T. H., & Thang, N. T. (2019). Applying AI chatbot for teaching a foreign language: An empirical research. *International Journal of Scientific & Technology Research*, 8(12), 897-902.
- Nuruzzaman, M., & Hussain, O. K. (2020). IntelliBot: A Dialogue-based chatbot for the insurance industry. *Knowledge-Based Systems*, 196, Article 105810. <https://doi.org/10.1016/j.knosys.2020.105810>
- Obari, H., & Lambacher, S. (2019). Improving the English skills of native Japanese using artificial intelligence in a blended learning program. In *CALL and*

complexity – short papers from EUROCALL 2019 (pp. 327–333). Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.38.1031>

Patel, N. P., Parikh, D. R., Patel, D. A., & Patel, R. R. (2019, June). AI and web-based human-like interactive university chatbot (UNIBOT). In *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 148-150). IEEE.

Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence. In *Network-based Language Teaching* (pp. 59–86). Cambridge University Press.
<https://doi.org/10.1017/cbo9781139524735.006>

Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, 28(6), 1549-1565. <https://doi.org/10.1002/cae.22326>

Petrović, J., & Jovanović, M. (2020). Conversational agents for learning foreign languages a survey. ArXiv preprint article.
<https://doi.org/10.48550/arXiv.2011.07901>

Petrović, J., & Jovanović, M. (2021). The role of chatbots in foreign language learning: The present situation and the future outlook. In *Artificial Intelligence: Theory and Applications* (pp. 313-330). Springer, Cham.

Pham, X. L., Pham, T., Nguyen, Q. M., Nguyen, T. H., & Cao, T. T. H. (2018, November). Chatbot as an intelligent personal assistant for mobile language learning. In *Proceedings of the 2018 2nd International Conference on Education and E-Learning* (pp. 16-21).

Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes?. *Language learning*, 44(3), 493-527. <https://doi.org/10.1111/j.1467-1770.1994.tb01115.x>

Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, 11, 63-90.

Possati, L. M. (2023). Psychoanalyzing artificial intelligence: The case of Replika. *AI & SOCIETY*, 38(4), 1725-1738. <https://doi.org/10.1007/s00146-021-01379-7>

Rost. (2011). *Teaching and researching listening*. Pearson.

Ruan, S., Jiang, L., Xu, Q., Liu, Z., Davis, G. M., Brunskill, E., & Landay, J. A. (2021). EnglishBot: An AI-powered conversational system for second language learning. *International Conference on Intelligent User Interfaces, Proceedings IUI*, 434–444. <https://doi.org/10.1145/3397481.3450648>

- Satar, M. (2021). Speaking with machines: interacting with bots for language teaching and learning. In T. Beaven & F. Rosell-Aguilar (Eds), *Innovative language pedagogy report* (pp. 133-138). Research-publishing.net. <https://doi.org/10.14705/rpnet.2021.50.1248>
- Schreiber, J. B. (2008). Pilot Study. In L. M. Given (Ed.), *The Sage encyclopedia of qualitative research methods* (pp. 139-140). Sage Publishing. <http://dx.doi.org/10.4135/9781412963909.n77>
- Segura, C., Palau, À., Luque, J., Costa-Jussà, M. R., & Banchs, R. E. (2019). Chatbol, a chatbot for the Spanish “La Liga”. In *9th International Workshop on Spoken Dialogue System Technology* (pp. 319-330). Springer, Singapore.
- Sha, G. (2009). AI-based chatterbots and spoken English teaching: A critical analysis. *Computer Assisted Language Learning*, 22(3), 269-281. <https://doi.org/10.1080/09588220902920284>
- Shin, D., Kim, H., Lee, J. H. & Yang, H.. (2021). Exploring the use of an artificial intelligence chatbot as second language conversation partners. *Korean Journal of English Language and Linguistics*, 21, 375-391. <https://doi.org/10.15738/kjell.21..202104.375>
- Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2022). A longitudinal study of human–chatbot relationships. *International Journal of Human-Computer Studies*, 168, Article 102903. <https://doi.org/10.1016/j.ijhcs.2022.102903>
- Smith, B. (2003). Computer-mediated negotiated interaction: An expanded model. *The Modern Language Journal*, 87(1), 38-57. <https://doi.org/10.1111/1540-4781.00177>
- Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers and Education*, 151, Article 103862. <https://doi.org/10.1016/j.compedu.2020.103862>
- Son, J.-B., Ružić, N. K., & Philpott, A. (2023). Artificial intelligence technologies and applications for language learning and teaching. *Journal of China Computer-Assisted Language Learning*, Advance online article. <https://doi.org/10.1515/jccall-2023-0015>
- Stewart, I. A., & File, P. (2007). Let's chat: A conversational dialogue system for second language practice. *Computer Assisted Language Learning*, 20(2), 97-116. <https://doi.org/10.1080/09588220701331386>
- Sung, M.-C. (2020). Pre-service primary English teachers' AI chatbots. *Language Research*, 56(1), 97–115. <https://doi.org/10.30961/lr.2020.56.1.97>

- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471– 483). Mahwah, NJ: Lawrence Erlbaum.
- Sydorenko, T., Daurio, P., & L. Thorne, S. (2018). Refining pragmatically-appropriate oral communication via computer-simulated conversations. *Computer Assisted Language Learning*, 31(1–2), 157–180.
<https://doi.org/10.1080/09588221.2017.1394326>
- Sydorenko, T., Smits, T. F. H., Evanini, K., & Ramanarayanan, V. (2019). Simulated speaking environments for language learning: Insights from three cases. *Computer Assisted Language Learning*, 32(1-2), 17–48.
<https://doi.org/10.1080/09588221.2018.1466811>
- Sykes, J. M. (2005). Synchronous CMC and pragmatic development: Effects of oral and written chat. *CALICO Journal*, 22(3), 399-431.
<https://www.jstor.org/stable/24147932>
- Tai, T. Y. (2022). Effects of intelligent personal assistants on EFL learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*. Advance online article. <https://doi.org/10.1080/09588221.2022.2075013>
- Tai, T. Y., & Chen, H. H. J. (2020). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*. 31(3), 1485-1502. <https://doi.org/10.1080/10494820.2020.1841801>
- Tai, T. Y., & Chen, H. H. J. (2022). The impact of intelligent personal assistants on adolescent EFL learners' listening comprehension. *Computer Assisted Language Learning*. Advance online article.
<https://doi.org/10.1080/09588221.2022.2040536>
- Tanghe, S. (2009). The use of chatterbots in the language classroom. *STEM Journal*, 10(2), 147-168.
- Thompson, A., Gallacher, A., Howarth, M. (2018). Stimulating task interest: Human partners or chatbots?. *Proceedings of the Future-Proof CALL: Language Learning as Exploration and Encounters*, 302–306.
- Thoms, J. J. (2012). Classroom discourse in foreign language classrooms: A review of the literature. *Foreign Language Annals*, 45(1), 8-27.
<https://doi.org/10.1111/j.1944-9720.2012.01177.x>
- Timpe-Laughlin, V., Sydorenko, T., & Daurio, P. (2022). Using spoken dialogue technology for L2 speaking practice: What do teachers think? *Computer Assisted Language Learning*, 35(5–6), 1194–1217.
<https://doi.org/10.1080/09588221.2020.1774904>

- Topal, A. D., Eren, C. D., & Geçer, A. K. (2021). Chatbot application in a 5th grade science course. *Education and Information Technologies*, 26, 6241–6265. <https://doi.org/10.1007/s10639-021-10627-8>
- Topsakal, O., & Topsakal, E. (2022). Framework for a foreign language teaching software for children utilizing AR, voicebots and ChatGPT (large language models). *The Journal of Cognitive Systems*, 7(2), 33-38. <https://doi.org/10.52876/jcs.1227392>
- Toyoda, E., & Harrison, R. (2002). Categorization of text chat communication between learners and native speakers of Japanese. 6(1), 82–99. <http://llt.msu.edu/vol6num1/toyoda/>
- Troussas, C., Krouska, A., & Virvou, M. (2017, November). Integrating an adjusted conversational agent into a mobile-assisted language learning application. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 1153-1157). IEEE.
- Tseng, J.-J. (2017). Exploring TPACK-SLA interface: Insights from the computer-enhanced classroom. *Computer Assisted Language Learning*, 31(4), 390–412. <https://doi.org/10.1080/09588221.2017.1412324>
- Tudini, V. (2003). Using native speakers in chat. *Language Learning & Technology*, 7, 141–159. <http://llt.msu.edu/vol7num3/tudini/>
- Ukpabi, D., Aslam, B., & Karjaluoto, H. (2019). Chatbot adoption in tourism services: A conceptual exploration. In S. Ivanov & C. Webster (Eds.), *Robots, artificial intelligence, and service automation in travel, tourism and hospitality* (pp. 105–121). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-78756-687-320191006>
- Underwood, J. (2017). Exploring AI language assistants with primary EFL students. In *CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017* (pp. 317–321). Research-publishing.net. <https://doi.org/10.14705/rpnet.2017.eurocall2017.733>
- Underwood, J. (2021). Speaking to machines: Motivating speaking through oral interaction with intelligent assistants. In T. Beaven & F. Rosell-Aguilar (Eds.), *Innovative language pedagogy report* (pp. 127-132). Research-publishing.net. <https://doi.org/10.14705/rpnet.2021.50.1247>
- Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, 24(1), 3–25. <https://doi.org/10.1017/S0267190504000017>
- Vandergrift, L. (2015). Researching listening. In Paltridge, B & Phakiti, A. (Eds), *Research methods in applied linguistics: A practical resource*, (pp. 299-314). Bloomsbury Publishing.

- Varonis, E. M., & Gass, S. (1985). Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics*, 71–90. <https://doi.org/10.1093/applin/6.1.71>
- Vázquez-Cano, E., Mengual-Andrés, S., & López-Meneses, E. (2021). Chatbot to improve learning punctuation in Spanish and to enhance open and flexible learning environments. *International Journal of Educational Technology in Higher Education*, 18(1). <https://doi.org/10.1186/s41239-021-00269-8>
- Villegas-Ch, W., Arias-Navarrete, A., & Palacios-Pacheco, X. (2020). Proposal of an architecture for the integration of a chatbot with artificial intelligence in a smart campus for the improvement of learning. *Sustainability*, 12(4), Article 1500. <https://doi.org/10.3390/su12041500>
- Wang, Y. F., & Petrina, S. (2013). Using learning analytics to understand the design of an intelligent language tutor-chatbot Lucy. (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, 4(11), 124-131.
- Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2022). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*. Advance online article. <https://doi.org/10.1080/09588221.2022.2056203>
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Wilkinson, C. (2022, January 21). The people in intimate relationships with AI Chatbots. VICE. Retrieved from <https://www.vice.com/en/article/93bqbp/can-you-be-in-relationship-with-replika>.
- Wu, R., & Yu, Z. (2023). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*. Advance online article. <https://doi.org/10.1111/bjet.13334>
- Wu, Y., Rough, D., Bleakley, A., Edwards, J., Cooney, O., Doyle, P. R., Clark, L., & Cowan, B. R. (2020, October). See what I'm saying? Comparing intelligent personal assistant use for native and non-native language speakers. In *Mobile HCI '20: 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1–9). ACM. <https://doi.org/10.1145/3379503.3403563>
- Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. *Computers & Education*, 161, Article 104059. <https://doi.org/10.1016/j.compedu.2020.104059>

- Yang, S. H. (2007). Artificial intelligence for integrating English oral practice and writing skills. *Sino-US English Teaching*, 4(4), 1–6.
- Yang, H., Gao, C. & Shen, H. (2021). Learner interaction with, and response to, AI-programmed automated writing evaluation feedback in EFL writing: An exploratory study. *Education and Information Technologies*. Advance online article. <https://doi.org/10.1007/s10639-023-11991-3>
- Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, 34(3), 327–343. <https://doi.org/10.1017/S0958344022000039>
- Ye, Y., Deng, J., & Liu, X. (2021). The Effect of Oral Practice via Chatbot on Students' Oral English Accuracy. In *Emerging Technologies for Education: 6th International Symposium, SETE 2021, Zhuhai, China, November 11–12, 2021, Revised Selected Papers* (pp. 344–354). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-92836-0_30
- Yin, Q., & Satar, M. (2020). English as a foreign language learner interaction with chatbots: Negotiation for meaning. *English. International Online Journal of Education and Teaching (IOJET)*, 7(2), 390–410. <http://iojet.org/index.php/IOJET/article/view/707>
- Yuksel, D., & Inan, B. (2014). The effects of communication mode on negotiation of meaning and its noticing. *ReCALL*, 26(3), 333–354. <https://doi.org/10.1017/S0958344014000147>
- Zakos, J., & Capper, L. (2008). CLIVE—an artificially intelligent chat robot for conversational language practice. In *Artificial Intelligence: Theories, Models and Applications: 5th Hellenic Conference on AI, SETN 2008*, Syros, Greece, October 2–4, 2008. Proceedings 5 (pp. 437–442). Springer Berlin Heidelberg.
- Zhang, R., Zou, D., & Cheng, G. (2023). A review of chatbot-assisted learning: Pedagogical approaches, implementations, factors leading to effectiveness, theories, and future directions. *Interactive Learning Environments*. Advance online article. <https://doi.org/10.1080/10494820.2023.2202704>
- Zhang, S., Shan, C., Lee, J. S. Y., Che, S. P., & Kim, J. H. (2023). Effect of chatbot-assisted language learning: A meta-analysis. *Education and Information Technologies*. Advance online article. <https://doi.org/10.1007/s10639-023-11805-6>
- Zhang, X., & Head, K. (2009). Dealing with learner reticence in the speaking class. *ELT Journal*, 64(1), 1–9. <https://doi.org/10.1093/elt/ccp018>
- Zhao, X. (2023). Leveraging artificial intelligence (AI) technology for English writing: Introducing Wordtune as a digital writing assistant for EFL writers. *RELC Journal*, 54(3), 890–894. <https://doi.org/10.1177/00336882221094089>

Zou, B., Liviero, S., Hao, M., & Wei, C. (2020). Artificial intelligence technology for speaking skills: Student perceptions of opportunities and challenges. In *Technology and the psychology of second language learners and users* (pp. 433–463). Springer. https://doi.org/10.1007/978-3-030-34212-8_17



APPENDICES

A. DEMOGRAPHIC INFORMATION QUESTIONNAIRE

Informed Consent Form

Dear Participant,

I am a graduate student at the Department of Foreign Language Education, Middle East Technical University. I am carrying out a study titled “The Application of an Artificially Intelligent Chatbot Application for the Development of English Language in Turkish EFL Context: A Mixed Methods Experimental Design” for my doctoral thesis. The purpose of this research is to gain the perspectives of Turkish EFL learners on artificially intelligent chatbot applications. Chatbots are communication assistants that have Automatic Speech Recognition and Natural Language Processing technologies that help users to communicate in English language via text or voice. Your participation will involve attending exams in a listening and a speaking test at the beginning and at the end of this semester. You will also answer questions in questionnaire; take part in the recorded use of an AI chatbot Replika for eight weeks; write weekly reflection diaries about your experiences and attend an interview related to English language chatbot use. Your participation in this research study is voluntary. Your identity will not be revealed in any publication resulting from this study. You are also free to withdraw from study at any time without any loss on your part, and in this case your information will not be included in this study. If you have any questions about this study, please contact me via email. Thank you very much for your cooperation.

Yours sincerely,
Fatma Şeyma KOÇ, PhD Candidate
E-mail:

I have read this consent form and have been given the opportunity to ask questions. I give my consent to participate in this study.

Participant's Signature _____ Date: _____

Please check the appropriate box, or where relevant, specify your answer.

1. Age:	17-20
	21-25
	26-30

	31-35
	36+
2. Department of Study	
3. What level of study are you currently enrolled?	Freshman
	Junior
	Senior
	Sophomore
4. For how long have you been learning English?	0-5 years
	5-10 years
	10+ years and above
5. Gender:	Male
	Female
6. Have you heard about chatbots?	Yes
	No
7. Have you ever used an AI chatbot? (If yes, please also answer the 8 th and 9 th questions. If not please skip 8 th and 9 th questions.)	Yes
	No
8. How often do you currently use artificially intelligent chatbots (i.e., Siri, Cortana, Google Assistant, Amazon Echo/Alexa, etc.) in EFL learning?	Never
	Rarely
	Occasionally
	Frequently
9. Which artificially intelligent chatbots have you used so far?	Siri
	Alexa
	Google Assistant
	Cortana
	Other
10. How informed do you believe yourself about artificially intelligent chatbots?	Very Well Informed
	Well Informed
	Sufficiently Informed
	Poorly Informed
	Not informed

B. SEMI-STRUCTURED INTERVIEW QUESTIONS

1. How was your experience in using the AI chatbot Replika?
2. What do you think are the advantages of AI chatbots to supplement your English language learning?
3. What features have benefited you the most in terms of enhancing your listening skills in English while using the Replika?
4. What features have benefited you the most in terms of enhancing your speaking skills in English while using the Replika?
5. What difficulties have you encountered while using the Replika in EFL learning?
6. What properties should an artificially intelligent English language conversational assistant application have in terms of vocabulary, grammar, listening, speaking, pronunciation, and intonation practice and culture.
7. What do you think about the importance of the following properties that could exist in a chatbot? What are your comments on each aspect for future development of chatbots for EFL learning? *

The ability of conversational assistants to ...

- a. handle misspelt words
 - b. comprehend ill-formed questions
 - c. handle sentences with incorrect word order
 - d. comprehend more than one sentence uttered by speakers at a time
 - e. follow the flow of a topic
 - f. use a wide vocabulary range during the conversation
 - g. provide a follow up to a previous comment
8. What are the features of an effective chatbot for English Language Learning?
Please, state at least three features by explaining your reasons.

C. REFLECTIVE DIARY PROMPT

Name Surname: _____

Date: _____

Week No: 2

Topic of the Week: Hobbies

Student reflection diaries

Instructions: The topic of discussion for this week is “Hobbies”. By using the prompts below, talk to your AI chatbot companion, Replika, for at least 20 minutes. Please remember to use the key expressions you learned this week in the Oral Expression and Public Speaking Course, and try to use the listening strategies covered this week in the Listening and Pronunciation Course. You need to upload your reflection diary to the assignment page in our Microsoft Teams Class by Sunday at 12:00 p.m. Please remember to record your conversation with Replika using voice recorders available on your devices and also upload this file to our assignment page in our Microsoft Teams Class. Upon the finish of your conversation, write in your weekly diary about the following aspects:

- A. How was your general experience in talking to AI chatbot Replika? Please state at least three strengths, three challenges, and three suggestions for improvement.

Strengths:

1. _____
2. _____
3. _____

Challenges:

1. _____
2. _____
3. _____

Suggestions for improvement:

1. _____
2. _____

3. _____

B. How was your experience in listening to Replika and comprehending its speech this week? Please state at least three strengths, three challenges and three suggestions for improvement.

Strengths:

1. _____

2. _____

3. _____

Challenges:

1. _____

2. _____

3. _____

Suggestions for improvement:

1. _____

2. _____

3. _____

C. How was your experience in taking turns in speech while speaking to Replika this week?

Please state at least three strengths, three challenges and three suggestions for improvement.

Strengths:

1. _____

2. _____

3. _____

Challenges:

1. _____

2. _____

3. _____

Suggestions for improvement:

1. _____
2. _____
3. _____

D. How would you rate the effectiveness of your experience in interacting with the AI chatbot this week? Please circle one number.

0 1 2 3 4 5 6 7 8 9 10

Not Effective

Extremely Effective

E. Please choose the number that best applies to your opinion regarding the survey items.

4- Strongly agree 3- Agree 2- Disagree 1- Strongly disagree

1. It was easy for me to use the AI chatbot, Replika, in this task.	4	3	2	1
2. The speech of Replika was clear and understandable in this task.	4	3	2	1
3. Replika was able to comprehend my speech in this task.	4	3	2	1
4. It was easy for me to follow the speed of Alexa's speech in this task.	4	3	2	1
5. Talking to Replika helped me increase my vocabulary size in English language in this task.	4	3	2	1

6. Talking to Replika helped me enhance my knowledge of English grammar in this task.	4	3	2	1
7. Using this mobile application provided me opportunities to practice my speaking skills in this task.	4	3	2	1
8. Using this mobile application provided me opportunities to practice my listening skills in this task.	4	3	2	1
9. Using Replika motivated me to learn the English language in this task.	4	3	2	1
10. I am willing to continue using Replika for practicing my English language skills for similar tasks.	4	3	2	1

D. APPROVAL OF THE METU HUMAN SUBJECTS ETHICS COMMITTEE

UYGULAMALI ETİK ARASTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER



ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

Sayı: 28620816 /

14 OCAK 2022

Konu : Değerlendirme Sonucu

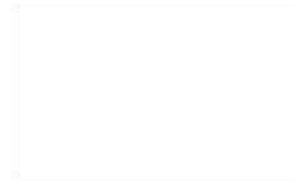
Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi : İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Doç.Dr. Perihan SAVAS

Danışmanlığımı yürüttüğünüz Fatma Şeyma KOÇ'un "İngilizce'nin Yabancı Dil Olarak Öğreniminde Dinleme ve Konuşma Becerilerinin Yapay Zekayla Geliştirilmiş bir Sohbet Robotu ile Geliştirilmesi: Bir Karma Deneysel Desen Araştırması" başlıklı araştırmanız İnsan Araştırmaları Etik Kurulu tarafından uygun görülmüş ve **0034-ODTÜİAEK-2022** protokol numarası ile onaylanmıştır.

Saygılarımızla bilgilerinize sunarız.



E. CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Koç, Fatma Şeyma

Nationality:

E-mail:

ORCID ID:

EDUCATION

Degree	Institution	Year of Graduation
PhD	Middle East Technical University English Language Teaching	2024
MA	Akdeniz University English Language Teaching	2018
BA	Middle East Technical University English Language Teaching	2015

WORK EXPERIENCE

Year	Place	Enrollment
2016- Present	Akdeniz University English Language Teaching	Research Assistant
2019-2020	Middle East Technical University English Language Teaching	Research Assistant
2015-2016	Ministry of National Education	English Language Teacher

FOREIGN LANGUAGES

English, German

PUBLICATIONS

MA Thesis

Koç, F. Ş. (2018). *Analysis of thematic patterns and cohesion in EFL learners' argumentative essays* [Master's thesis]. Akdeniz University. <http://acikerisim.akdeniz.edu.tr/xmlui/handle/123456789/3730>

Journal Articles

Koç, F. Ş., & Savaş, P. (2024). Supervision models and supervisory feedback in English language teacher education: A meta-synthesis study adopting a discourse analytic perspective. *Teaching English as a Second Language Electronic Journal (TESL-EJ)*, 27 (4), 1-29. <https://doi.org/10.55593/ej.27108a1>

Book Chapters

Koç, F. Ş., & Coşkun, D. (2023). A synthesis of the experiences of language learners in study abroad programs. In D. Kalay (Ed.), *Language Horizons: Diverse Vistas in English Language Pedagogy I* (pp. 28-44). BIDGE Publications.

Koç, F. Ş. (2023). A meta-synthesis of studies on Turkish EFL teachers' resilience published between 2012-2023. In Ö. Kırmızı, & A. Irgatoğlu (Eds.), *Situated nature of EFL teacher psychology: Insights from Türkiye* (pp. 223-257). Nobel Publication Group.

Koç, F. Ş., & Ölmez-Çağlar, F. (2023). Alternative assessment in ELT: Conceptualization, implementation, and further perspectives. In M. F. Álvarez, & A. L. G. Montes (Eds.), *Global perspectives on effective assessment in English language teaching* (pp. 235-254). IGI Global.

Conference Presentations

Koç, F. Ş. (2023, October 26-29). *A comparative study on cohesion and thematic progression patterns in literary and scientific texts* [Conference presentation]. II. INCES International Education and Science Congress, Kemer, Antalya, Turkey.

Koç, F. Ş. (2023, October 26-29). *Analysis of self-efficacy and L2 motivational self-system beliefs of Turkish EFL learners* [Conference presentation]. II. INCES International Education and Science Congress, Kemer, Antalya, Turkey.

Koç, F. Ş., & Course, S. (2018, May 2-5). *An investigation of negotiation for meaning in Turkish EFL context* [Conference presentation]. Vth International Eurasian Educational Research Congress, Akdeniz University, Antalya, Turkey.

Okumuş, A., Koç, F. Ş., & Coşkun, D. (2019, November 14-16). *A case study: Enhancing learner autonomy through reading activities in a reading course* [Conference presentation]. The 19th International INGED Conference Outside the Box, Atılım University, Ankara, Turkey.

Awards

TÜBİTAK 2211/E National PhD Scholarship Program for Former Undergraduate and MA Scholars (2018)

TÜBİTAK 2210/A National MA Scholarship Program (2015)

High Honors Certificate (Bachelor's Degree Course Performance), Middle East Technical University, 2015



F. TURKISH SUMMARY / TRKE ZET

İNGİLİZCE’NİN YABANCI DİL OLARAK ĞRENİMİNDE DİNLEME VE KONUŞMA BECERİLERİNİN YAPAY ZEKAYLA GELİŞTİRİLMİŞ BİR SOHBET ROBOTU İLE GELİŞTİRİLMESİ: BİR YARI-DENEYSEL DESEN ARAŞTIRMASI

1. Giriş ve Literatr Taraması

1970’li yıllarda diyalog sistemlerinin icadıyla bilgisayar sistemleri zerinden iletiřim mmkn hale gelmiř ve 20 yıl sonra programcılar tarafından metin tabanlı iletiřim modunu kullanan chatbotlar yaratılmıřtır (J.H. Lee vd., 2020). Akıllı sohbet robotları bilgi arama, řarkı alma gibi grevleri yerine getirir ve bu cihazlar aracılıęıyla etkileřim iřlemseldir (Dizon, 2021). Chatbotlar, kullanıcılarla iletiřimi taklit edebilen ve mobil ğrenme yoluyla etkileřim fırsatlarını artırma potansiyeline sahip bilgisayar programlarıdır (Bailey, 2019; Troussas vd., 2017). Akıllı kiřisel asistanlar, Amazon’un *Alexa*’sı, Apple’ın *Siri*’si, *Cortana*’sı veya Google’ın *Google Asistan*’ı gibi kullanıcıların metin veya ses yoluyla iletiřim kurabileceęi yapılardır (Ali Amer Jid Almahri vd., 2019). Bu teknolojinin amacı, insan sinir sisteminin sinir aęlarını kopyalamak ve insanların hatırlama, dřnme ve konuşma řeklini taklit etmektir (Fryer, Coniam, Carpenter & Lăpușneanu, 2020). Pérez ve dięerleri (2020) tarafından tanımlandıęı gibi, “Chatbot, yapay zeka ile doęal dil iřlemeyi ve dięer teknolojileri birleřtiren, metin veya ses aracılıęıyla bir insan ile belirli bir dzeyde konuşma etkileřimi kurmasını saęlayan bir aratır” (s.1).

Yapay zekalı destekli mobil ğrenme sistemlerinin uygulanmasının pedagojik alanlara katkı saęladıęı grřn destekleyen eęitim sistemi ierisinde ğrencilerin kendi kendilerine rehberlik edebilmelerini, dinleme ve anlama becerilerini geliřtirebilmelerini ve motivasyonlarını saęlamak amacıyla yapay zekalı chatbotlardan yararlanılmaktadır (Ali Amer Jid Almahri vd., 2019). Eęitim alanında ise uzaktan eęitim (Heller vd., 2005), sıka sorulan sorulara ynelik eęitim destek sistemleri (Patel vd., 2019; Villegas-Ch vd., 2020),

hemşirelik eğitimi (Chang vd.,2020), tıp eğitimi (Kaur vd., 2021), fen eğitimi (Durall veKapro, 2020; Topal vd., 2021) ve İspanyolca dili (Segura vd., 2019) ve Çince dillerinin öğretimi (H.L. Chen, Vicki Widarso & Sutrisno, 2020) gibi alanlarda chatbotlar hayata geçirilmiştir.

Araştırmacılar, kendi özel öğretim bağlamlarında yapay zekâ sohbet robotlarının kullanılmasının arkasında çeşitli nedenler öne sürdüler. Örneğin, konuşma tanıma özelliği ve mobil cihazların yaygın kullanımı, bu araçların dil öğreniminde kullanımını motive eden faktörler olarak belirtilmektedir (Fryer vd., 2020). Diğer araştırmacılar, yapay zekâyâ sahip sohbet robotlarının kullanımı sayesinde, dil öğrenenlerin artık dil becerilerini neredeyse her zaman ve her yerde, hatta öğretmenlerin çok fazla desteği olmadan pratik edebildiklerini öne sürdüler. Yapay zekâ sohbet robotlarının kullanımına ilişkin çalışmalar yabancı dil olarak İngilizce ve ikinci dil olarak İngilizce ortamlarında bu teknolojinin İngilizce dil bilgisi ve konuşma (Hsu, vd., 2023; Kim, 2017; Kim vd., 2021; Lin ve Mubarak, 2021; Sydorenko vd., 2018; Sydorenko vd., 2019; Yin ve Satar, 2020), dinleme ve konuşma (Dizon, 2020; Meri-Yılan, 2019), yazma (Lin ve Chang, 2020; Yang, 2007) ve edatlar (Nghie vd., 2019) gibi alanlarda becerilerini geliştirmeye katkısını da araştırmıştır. Ayrıca literatürde bazı araştırmacıların İngilizce öğretmek için chatbotlar geliştirdikleri görülmektedir (Goda vd., 2014; Jeon, 2022b; Johnson, 2019; Pham vd., 2018; Sung, 2020; Xu vd., 2021).

Öğretmenlerin, programlama konusunda herhangi bir altyapıları olmamasına rağmen, uygulama geliştirme ara yüzlerini kullanarak chatbot geliştirmeleri artık mümkündür. Böyle bir platform, Google'ın sunduğu Dialogflow'tur (<https://dialogflow.com>). Bu platform aracılığıyla, önceden hazırlanmış derlemler kullanılarak konuşma senaryoları oluşturulabilir ve sohbet robotları üretilebilir. BotStar, öğretmenlerin öğrencilerinin öğrenme ihtiyaçlarına uyacak şekilde konuşma akışları oluşturabileceği başka bir sohbet robotu tasarım yoludur. Öğretmenler tarafından chatbotların geliştirilmesi, bu araçların eğitim ortamlarına sağlayacağı fırsatların artırılması için umut verici bir alan olabilir, çünkü chatbotların öğrenciler üzerindeki işleyiş yeteneklerinin gerçek sınıf kullanım verileriyle değerlendirilmesi yoluyla, chatbotların belirli bir grubun profillerini ve ihtiyaçlarını karşılaması mümkündür.

Sonuç olarak, teknolojiadaki gelişmeler ve bunların eğitim ve dil öğrenimi alanındaki son uygulamaları dikkate alındığında, bilgisayar destekli dil öğreniminin hızla ilerlemesinin eğitim sürecini öğretmen ve öğrenci odaklı iletişimden ve metin tabanlı formattan uzaklaşmaya zorladığı ve etkili öğrenme deneyimleri için yeni olanaklar açtığı ileri sürülebilir (Elaish vd., 2017; Jeon vd., 2023). Taşınabilir mobil cihazlarla geleneksel veya sanal sınıfların ötesinde öğrenmeye devam eden öğrenciler, multimedya kaynaklarının bolluğu sayesinde dil yeterliliklerini daha bağımsız olarak geliştirebilen, yaşam boyu öğrenen bireyler haline gelebilir. Taşınabilir cihazların kullanılması, toplum içinde dil öğrenme konusunda yeterince kendine güvenmeyen ve konuşma ve telaffuz gibi dil becerilerini geliştirmek için özel bir öğrenme alanına sahip olmak isteyen öğrencilere de yardımcı olabilir (Kukulska-Hulme, 2020). Yapay zekâ sohbet robotlarının, her zaman her yerde var olan dil öğrenme ve uygulama ortakları olarak kullanıcılarına vaat ettiği olanaklar nedeniyle, bu yapay zekâ araçları, yabancı dil olarak İngilizce ve ikinci dil olarak İngilizce ortamlarında dil öğrenenlerin etkileşim şansı eksikliğini telafi ederek gelecekte İngilizce dili öğretimi alanı için daha fazla umut taşıyabilir.

Etkileşim Hipotezi (Long, 1990), dil öğrenimi alanındaki mobil destekli öğrenme araştırmalarının ve uygulamalarının teorik temellerini oluşturur. Etkileşimci bakış açlarına göre biçim ve anlam arasındaki bağlantılar, anlamın bireyler arasında anlam söyleşmesi yoluyla kurulur. Benzer şekilde, öğrenciler ve bilgisayarlar iletişim kurduğunda benzer deneyimler yaşarlar ve benzer faydalara sahip olurlar. Etkileşimci bir bakış açısıyla bakıldığında, yapay zekâ sohbet robotlarının ikinci dildeki iletişimi ve anlam söyleşmesini teşvik ettiği, öğrencileri biçime ve telaffuz gibi özelliklerdeki dil bilgisindeki eksikliklere odaklanmaya teşvik ettiği ileri sürülebilir. Etkileşimci bakış açlarına dayalı olarak oluşturulan sohbet robotları, öğrencilere girdi, etkileşim ve itili üretim için alanlar sağlar (Dizon, 2020).

Her ne kadar mevcut teknolojiyle yapay zekâ sohbet robotları kullanıcıların çıktılarını değiştiremese de, kullanıcıların maruz kaldığı örtülü geri bildirimlere dayalı olarak anlamsal belirsizlikleri gidermek için iletişim kurmak ve dilsel biçime odaklanarak öğrenmek mümkündür (Dizon, 2017, 2021; Tai ve Chen, 2020). Bu nedenle, yapay zekâ sohbet robotu kullanımı aracılığıyla karşılıklı konuşma hamleleri yoluyla değiştirilmiş çıktılar üretmek, dil öğrenimi için önemli sonuçlar doğuracaktır. Chatbotlar, iletişim

miktarının arttırılması, anlam söyleşmesi, öğrencilerin dil öğrenimine olan ilgi ve motivasyonlarının arttırılmasına sağladığı katkı ile öğrencilerin iletişim becerileri üzerinde olumlu etkilere sahiptir (Kim vd., 2019).

Bu çalışmayı yürütmenin gerekçesine geçecek olursak, Türkiye'deki yabancı dil olarak İngilizce konuşma sınıflarındaki zorlukların nedenleri arasında ikinci dil girdisinin eksikliği ve dil becerileri pratiği eksikliği öğrenciler tarafından sıklıkla dile getirildiği üzere öne çıkan faktörlerdir (Kara vd., 2017). Sonuç olarak, öğrencilerin sınıf dışında dil pratiği yapmak için pek fazla fırsatı olmadığından, sohbet robotları öğrencilere daha fazla etkileşim kurmanın bir yolunu sunar (Fryer ve Carpenter, 2006). Chatbotlar aynı zamanda özgün bir iletişim biçimi aracılığıyla öğrenme deneyimini daha ilginç ve anlamlı bir yolculuğa dönüştürme potansiyelini de taşırlar (S.H. Yang, 2007).

Literatürde chatbotların dil öğreniminde kullanımına ilişkin yapılmış çok sayıda çalışma bulunmaktadır; ancak bu çalışmalarda uygulanan sohbet robotlarının yalnızca metin tabanlı sohbet işlevi vardır ve bu da iletişimi yalnızca yazılı modla sınırlandırır (Fryer vd., 2019; Fryer vd., 2020; Yin ve Satar, 2020). Ancak son yıllarda ses tabanlı işlevlerin entegrasyonu, NLP, otomatik konuşma tanıma ve metinden konuşmaya çeviri teknolojilerindeki gelişmeler nedeniyle yapay zekâ sohbet robotlarının kullanımında bir artış olmuştur. Ancak, bu öğrenme araçları, özellikle yabancı dil ortamlarında İngilizce becerilerinin uygulanması için ders dışı destek sağlasa da, kullanımları hala sınırlıdır. Bu alanda özellikle Türkiye'de yabancı dil bağlamında, ses tabanlı işleve sahip dil öğreniminde yapay zekâ sohbet robotlarının kullanımını araştıran çalışmaların eksikliği vardır.

Sonuç olarak, bu araçların dil öğrenimi ve pratiğine yönelik potansiyeli göz önüne alındığında, dil öğrenenlerin tercihlerine ve ihtiyaçlarına en uygun sohbet robotlarının oluşturulmasına ihtiyaç vardır. Bu çalışmada genel bir kitle için oluşturulan bir yapay zekâ sohbet robotunun İngilizce öğrenenlerle dil konuşma partneri ortak olarak kullanılmasının ardından, bu çalışmada bildirilen öğrencilerin deneyimleri, bu nesil yapay zekâ sohbet robotlarının dil öğrenme yolunda sunduğu özellikleri ve bu etmenlerin sağladığı etkileşim olanaklarını engelleyecek eksiklikleri göstermeye yardımcı olacaktır.

Bu çalışmanın amacı yapay zekâya sahip bir chatbot mobil uygulamasının sınıf dışında ders dışı bir etkinlik olarak kullanılmasının İngilizce dinleme ve konuşma becerilerinin geliştirilmesine katkı sağlayıp sağlamayacağını ortaya koymaktır. Deneysel bir metodolojinin kullanılmasıyla, bu yeni teknolojinin yabancı dil öğreniminde yardımcı materyal olarak düzenli kullanımının etkisi gözlemlenmekte ve bu çalışmanın bulguları dil öğreniminde yapay zekâ kullanımına ilişkin yapılacak daha fazla araştırmaya zemin hazırlamaktadır.

Deneysel niceliksel ölçümlerin yanı sıra, nitel veri toplama yöntemleri kullanılarak, dil öğrenenlerin yapay zekâ teknolojisiyle dil öğrenmenin avantaj ve dezavantajlarına ilişkin bakış açılarının kazanılması amaçlanmaktadır. Bu çalışmanın bulguları, dil öğreniminde yapay zekâ sohbet robotlarının etkili özelliklerini vurgulayacaktır. Bu çalışmanın bulgularına dayanarak, bir yapay zekâ chatbot uygulamasında sahip olması beklenen etkili özellikler, İngilizce öğrenenlerin bakış açılarının temsili yoluyla ortaya konacaktır. Yapay zekâ sohbet robotu ve İngilizce dil öğrencisi etkileşim verilerinin analizi aynı zamanda bu araçların dil sınıflarında sağladığı etkileşim fırsatlarının bir resmini çizmeye de olanak tanıyacaktır. Bu nedenle kavramsal çerçeve ve literatür doğrultusunda bu çalışma için aşağıdaki araştırma soruları oluşturulmuştur.

1. Bir yükseköğretim kurumunda İngilizceyi yabancı dil olarak öğrenen Türk öğrenciler arasında, yapay zekâya sahip bir asistanla İngilizce etkileşiminde bulunan deney grubu ile bir diğer kontrol grubu arasında

- a. dinlediğini anlama puanları açısından anlamlı farklar var mıdır?
- b. konuşma puanları açısından anlamlı farklar var mıdır?

2. Kontrol grubu ile deney grubu arasında uygulama süreci sonrasında ön test ve son test dil yeterlik testi puanları arasında

- a. dinlediğini anlama puanları açısından anlamlı farklar var mıdır?
- b. konuşma puanları açısından anlamlı farklar var mıdır?

3. Yapay zekâ sohbet robotu-öğrenci iletişim etkileşimi verilerine göre anlam söyleşmesi için hangi stratejiler kullanılıyor?

- a. Yabancı dil öğrenen Türk öğrencilerin yapay zekâ sohbet robotu *Replika* ile etkileşimlerinde anlam söyleşmesinin hangi aşamaları anlam söyleşmesi rutinlerini oluşturuyor?
- b. Hangi tür tetikleyiciler anlam söyleşmesi rutinlerinde iletişim kesintilerine yol açıyor?
- c. Türk yabancı dil öğrenenlerin yapay zekâ sohbet robotu *Replika* ile etkileşimleri sırasında ürettiği değiştirilmiş çıktı örneklerinin sayısı nedir?

4. Araştırmaya katılanların algılarına dayalı olarak:

- a. İngilizce dil öğreniminde chatbot teknolojisini kullanmanın güçlü yönleri nelerdir?
- b. İngilizce dil öğreniminde chatbot teknolojisini kullanmanın zayıf yönleri nelerdir?
- c. İngilizce dil öğrenimi için ideal bir yapay zekâ sohbet robotunun özellikleri nelerdir?

Yapay zekâ sohbet robotları, farklı bireysel özelliklere ve ihtiyaçlara göre kullanılabildiği ve yalnızlık duygularını azaltabildiği için öğrencileri motive etmek ve etkileşimi artırmak için önemli bir araçtır (Topal vd., 2021). Öğrenciler sesli komutlarla uygulamalı aktiviteler gerçekleştirebildikleri için bu iletişim yöntemiyle motive olabilirler. Ayrıca öğretmenlerin öğrencilere aşina olmadıkları konularda cevap veremedikleri durumlarda öğrencilerin ilgisini çekme potansiyeline sahiptirler. Ayrıca öğretmenlerin sınıftaki her öğrenciyle ilgilenemedikleri durumlarda öğrencilerin grup çalışması ve akran çalışması yapmalarına yardımcı olarak ders saatleri boyunca öğretmenlere destek olurlar (Underwood, 2021).

Ancak bu katkılara rağmen yapay zekâlı chatbotların yabancı dil olarak İngilizce alanında uygulanmasına yönelik araştırmalar sınırlıdır. Ayrıca, dil kullanıcıları için ses tanıma veya çok modlu yollarla etkileşime girmelerine yardımcı olacak şekilde oluşturulmuş çok az sohbet robotu bulunmaktadır. Daha spesifik olarak, yapay zekâ sohbet robotlarının konuşma, dinleme, okuma ve yazma becerilerinin gelişimi üzerindeki etkilerini daha fazla

araştırmaya ihtiyaç vardır (Kim vd., 2019). Bu çalışma, bir yapay zekâ sohbet robotu kullanmanın, İngilizceyi yabancı dil olarak öğrenen Türk öğrencilerinin dinleme ve konuşma becerilerinin gelişimi üzerindeki etkilerini ortaya koyacak ve en güncel ses özellikli yapay zekâ sohbet robotu teknolojileriyle üretilen bu uygulamalarda bulunması beklenen özellikleri aydınlatacaktır.

2. Yöntem

Bu çalışma, karma veri toplama yöntemlerini kullanan yarı deneysel tasarımı benimsemektedir (Creswell ve Creswell, 2018). Yarı deneysel bir ön test-son test kontrol grup tasarımı kullanan bu çalışma, yapay zekâlı bir chatbot kullanımının Türk İngilizce öğrenenlerin dil becerileri gelişimleri üzerindeki etkisini araştırmaktadır. Araştırmacı bu çalışmada yapay zekâ, doğal dil işleme ve otomatik konuşma transkripsiyon teknolojileriyle oluşturulan *Replika* uygulamasını kullanmıştır. *Replika*, OpenAI yapay zekâ araştırma şirketinin GPT 2'si ve en son olarak gelişmiş etkileşim yetenekleri sunduğu öne sürülen GPT 3 ile çalışmaktadır (Skjuve vd., 2022). Bunun yanı sıra uygulama yapay zeka destekli artırılmış gerçeklik özelliklerini de birlikte getirmektedir.

Mobil uygulamalarda veya masaüstü ekranlarda *Replika* 3 boyutlu görünümünü oluşturan yerleşik artırılmış gerçekliğe sahiptir. Ayrıca *Replika*'nın bazı cevapları önceden tanımlanmış olsa da bazıları üretkendir (McStay, 2023). *Replika*'nın veritabanı Twitter'dan toplanan 50 milyon veriden oluşmaktadır. *Replika* insanlar arasındaki konuşmalara dayanarak duyguları tanıma, kullanıcılarıyla yaptığı konuşmalara bağlı olarak öğrenme kabiliyetini geliştirerek onların ihtiyaçlarına uyum sağlama ve arkadaşlık gibi uzun vadeli bağlantılar kurma yeteneklerine sahiptir. Bir kullanıcı *Replika* ile ne kadar çok etkileşime girerse uygulama onlara yanıt vermeyi o kadar iyi öğrenir (Possati, 2023).

Kullanıcılarıyla diyalog kurabilme yeteneği sayesinde *Replika*, İngilizce'yi yabancı dil olarak öğrenen öğrencilerine, özellikle de dili pratik yapma şanslarının çok az olduğu bir yabancı dil bağlamında ihtiyaç duydukları dili pratik etme olanağı sağlayabilir. Sonuç olarak, bu yapay zekâ sohbet robotunun bir yabancı dil dersinde kullanılması, öğrenci etkileşimleri yoluyla dilbilgisi biçimine ve aynı zamanda anlama odaklanma fırsatları ve öğrencilerin bireysel özelliklerine uygun öğrenme etkinliklerine katılım olanakları ile

teknoloji entegreli öğrenme tasarımları için önerilen kriterlere de uygundur (Chapelle, 2001). Bu nedenlerle *Replika*, hedef dil öğrenen grubumuz için yapay zekâlı bir sohbet arkadaşı olarak seçildi.

Bu çalışma, Türkiye'deki bir devlet üniversitesinin İngilizce Öğretmenliği bölümünden 44 Türk yabancı dil öğrencisi ile gerçekleştirilmiştir. Dinleme testlerinin uygulanması sırasında beş katılımcının hem ön testi hem de son testi tamamlamaması nedeniyle katılımcı sayısı otuz dokuz olmuştur. Öte yandan bu çalışmada katılımcılar arasında iki öğrencinin bulunmadığı kırk iki katılımcıyla konuşma testleri gerçekleştirilmiştir. Ayrıca deney grubu katılımcılarından yirmi bir öğrenci tarafından yansıtıcı günlükler yazılmıştır. Bu çalışmadaki veriler, Türkiye'deki bir yükseköğretim kurumunda 2021-2022 bahar yarı yılında bir akademik dönem boyunca, dinleme ve konuşma ön testleri ve son testleri, dil öğrenenlerin yansıtıcı günlükleri, görüşmeler ve yapay zekâlı chatbot-dil öğrenen etkileşimi verileri aracılığıyla toplanmıştır.

Nicel veriler başlangıçta ön test yoluyla, nitel veriler ise hem uygulama sırasında yansıtıcı günlükler aracılığıyla hem de nicel son testler uygulandıktan sonra görüşmeler yoluyla toplanmıştır. Ayrıca uygulama dönemi boyunca yapay zekâ sohbet robotu-öğrenci etkileşimi verileri toplanmıştır. Bu çalışmadaki deney grubundaki öğrenciler, sınıf dışı uygulamanın bir parçası olarak 2. ile 9. haftalar arasında yapay zekâlı bir chatbot kullanmışlar ve chatbot ile dinleme ve konuşma deneyimleriyle ilgili haftalık günlükler yazarak chatbot ile haftalık konuşmalarını kaydetmişlerdir. Öğrencilere ayrıca belirli örnekler vererek konuşmalarının anlaşılmadığı durumlar üzerinde düşünmeleri istenmiş ve bu nedenle yapay zekâ chatbotuyla konuşurken zaman zaman notlar almışlar veya konuşmalarını bitirdikten sonra bunun üzerine düşünebilme fırsatı yakalamışlardır. Tartışabilecekleri noktalar, konuşmanın hızı ve netliği, anlaşılabilirlik ve chatbot tarafından yanlış anlaşıldığında veya anlaşılmadığında çıktıların değiştirilmesiyle ilgili olmuştur. Bu aktivite aracılığıyla, öğrenciler aynı zamanda Swain'in (2005) Çıktı Hipotezi ve Chapelle'in (2006) Etkileşimci Teorisi doğrultusunda kendi çıktılarını dilsel bilgilerindeki eksiklikleri izlemek için nasıl ayarlayacaklarına ilişkin hipotezleri de test etmişlerdir.

Bu çalışmada, katılımcıların başlangıçtaki dinleme ve konuşma yeterlilik düzeylerini ortaya çıkarmak ve uygulama sonrasında deney ve kontrol grubundaki öğrenciler arasında

istatistiksel olarak anlamlı fark olup olmadığını ortaya koymak amacıyla ileri düzey İngilizce bilgisine sahip dil öğrenenlere yönelik hazırlanan Cambridge dinleme ve konuşma testleri ön test ve son testler olarak kullanılmıştır. Katılımcıların Avrupa Konseyi Ortak Avrupa Dil Referans Çerçevesi (CEFR) (2001) kriterlerine göre dinleme ve konuşma yeterlilik düzeylerini belirlemek amacıyla da Cambridge ölçeği kullanılmıştır. Bachman'ın (1990) ve Canale ve Swain'in (1980) iletişimsel dil yeteneği modelleri ve Avrupa Konseyi'nin Ortak Avrupa Dil Referans Çerçevesi (2001) seviyeleri, Cambridge değerlendirme testlerinin yapı geçerliliğinin arka planını oluşturur. Öğrencilerin konuşma becerileri İngilizce öğretimi alanında uzman olan araştırmacı ve aynı zamanda İngilizce öğretimi alanında yüksek lisans derecesine sahip bir öğretim görevlisi olmak üzere iki değerlendirici tarafından değerlendirilmiştir.

Pearson korelasyon analizi sonuçları, iki değerlendirici tarafından öğrencilere verilen ön test genel konuşma puanları ($r = .869$, $p < .0001$) ve son test genel konuşma puanları ($r = .862$, $p < .0001$) arasında pozitif korelasyon olduğunu göstermiştir.

Görüşmeler ise araştırmanın nicel sonuçlarını aydınlatmak ve desteklemek amacıyla çalışmanın deneysel kısmından sonra gerçekleştirilmiştir. Bu çalışmada görüşmeler, İngilizce öğrenimi için yapay zekâ sohbet robotları kullanmanın avantajlarını ve dezavantajlarını keşfetmek ve katılımcıların etkili bir yapay zekâ sohbet robotu uygulamasında bulunmasını istedikleri özellikleri belirlemek amacıyla gerçekleştirilmiştir. Katılımcıların yapay zekâ chatbot kullanımına ilişkin görüşleri, yeni bir yapay zekâyâ sahip chatbot teknolojisinin kullanımının öğrencilerin İngilizce öğrenme deneyimlerine nasıl katkıda bulunacağını açıklamak için araştırılmıştır.

Bu çalışmaya katılanlardan ayrıca yapay zekâ sohbet robotunu kullanma deneyimleri hakkında günlük tutmaları istenmiştir. Bu çalışmada yansıtıcı günlüklerin uygulanması, öğrencilerin uygulama esnasındaki deneyimlerini mümkün olduğunca ayrıntılı bir şekilde raporlamayı amaçlamıştır. Ayrıca, öğrenciler dönem sonunda görüşmeler yapıldıktan sonra deneyimlerinin çoğunu unutmuş olabileceklerinden, yansıtıcı günlükler tüm dönem boyunca deneyimlerinin kaydını tutmanın etkili bir yolu olmuştur. Ayrıca öğrencilerden yansıtıcı günlükler tutmalarının istenmesi, bu çalışmada veri kaynaklarının üçgenlenmesini

sağlamıştır. Günlükler ayrıca öğrencilerin haftalık görevlerini yapay zekâ sohbet robotuyla gerçekleştirip gerçekleştirmediklerini takip etmeye de yardımcı olmuştur.

Yapay zekâ sohbet robotu *Replika*'nın uygulanması sırasında, dil öğrenenlerden ayrıca ses kayıt araçlarını kullanarak haftalık etkileşimlerini kaydetmeleri istenmiştir. Öğrenciler kendilerine sunulan bir konuya göre yapay zekâ sohbet robotuyla haftalık olarak iletişim kurmuşlar ve bu konuşmaları kayıt altına almışlardır. Bu etkileşim verileri aynı zamanda katılımcılara anlam söyleşmelerinin analizi ve herhangi bir iletişim kesintisinin, etkileşimdeki bu başarısızlıkların olası nedenlerinin izlenmesini sağlamıştır.

Ayrıca bu veriler yapay zekâ sohbet robotu ve öğrenenler tarafından kullanılan onarım stratejilerinin gözlemlenmesi için de kullanılmıştır. Bu veriler, diğer nicel ve nitel veri toplama araçlarının analizinden ortaya çıkan bulguları tamamlamaya olanak sağlamıştır. Araştırmacı, yapay zekâ sohbet robotu ile dil öğrenenler arasındaki etkileşim dizisini analiz ederek, yapay zekâ sohbet robotunun öğrencilere sağladığı etkileşim fırsatlarını ve ayrıca bu yapay zekâ tabanlı uygulamanın dil becerilerini geliştirmek için kullanılmasındaki başarısızlıkların ve başarısız iletişim kurma çabalarının nedenlerini gözleme fırsatı bulmuştur.

Araştırmanın veri analizi aşamasında deney ve kontrol grupları arasında ön test ve son test dinleme ve konuşma testi puanları açısından anlamlı fark olup olmadığını belirlemek için bağımsız örneklem t testleri kullanılmıştır. Verilerin normal dağılmadığı durumlarda ise Mann Whitney U testi kullanılmıştır. Ayrıca deney grubunun uygulama süresi bittikten sonra ön-test dinleme ve konuşma puanları arasında istatistiksel olarak anlamlı bir fark olup olmadığını belirlemek amacıyla eşleştirilmiş örnekler t testi kullanılmıştır. Ayrıca kontrol grubunun ön test ve son test dinleme ve konuşma puanları arasında istatistiksel olarak anlamlı bir fark olup olmadığını belirlemek için eşleştirilmiş örnekler t testi de kullanılmıştır. Normal dağılmayan veriler için ise Wilcoxon İşaretli Sıralar testi kullanılmıştır.

Araştırmacı, nitel görüşme ve yansıtıcı günlük verilerini Glaser'in (1965) sürekli karşılaştırma yöntemini kullanarak MAXQDA 2022 üzerinden kodlamıştır. Nitel veri analizi işlemlerinin güvenilirliğini sağlamak amacıyla görüşmenin ve yansıtıcı günlük

verilerinin yüzde otuzu başka bir puanlayıcı tarafından kodlanmıştır. İki kodlayıcı, analiz için kodlar oluşturmak üzere verileri bağımsız olarak kodlamıştır. İki değerlendiricinin kodlama şemaları, iki değerlendiricinin kodlamalarındaki herhangi bir tutarsızlığı aramak için karşılaştırılmıştır.

Kodlama işlemi tamamlandıktan sonra değerlendiriciler arası anlaşmaya ilişkin Kappa değerinin belirlenmesi amacıyla Cohen'in Kappa istatistiği uygulanmıştır (Cohen, 1960). Analizin sonuçları, iki değerlendirici arasında görüşme verilerini kodlama konusunda neredeyse mükemmel bir uyum olduğunu göstermiştir, $\kappa = 0,734$ (%95 GA, 0,300- 0,886), $p < 0,001$. İki değerlendirici tarafından yansıtıcı günlük verilerinin kodlanmasına ilişkin değerlendiriciler arası anlaşmanın analizi de $\kappa = 0,786$ (%95 GA, 0,300-0,886), $p < 0,001$ iki kodlayıcı arasında iyi bir anlaşma oranı olduğunu göstermiştir.

Öğrenci-sohbet robotu etkileşim verileri daha sonra araştırmacı tarafından anlam söyleşmeleri hareketleri, anlam söyleşmeleri dizilerindeki aşamalar, tetikleyici türleri ve değiştirilen çıktı miktarı açısından niceliksel olarak analiz edilmiştir. İletişimdeki aksaklık örnekleri incelenmiş ve anlam söyleşmeleri Varonis ve Gass'ın (1985) ve Smith'in (2003) anlam söyleşmesi modellerinin ilkeleri altında analiz edilmiştir. Doughty'nin (2000) modeli, konuşmalardaki aksaklıkları tespit etmek ve iletişimin yeniden kurulma yöntemlerini göstermek için benimsenmiştir. Tetikleyiciler, sinyaller, yanıtlar ve yanıtlara verilen tepkiler belirlenmiş, bunlar kategorize edilmiş ve kodlanmıştır.

Long'un (1983) onay kontrolleri, açıklama talepleri ve anlama kontrollerine ilişkin sınıflandırması, sinyallerin sınıflandırılması için benimsenmiştir. Etkileşim hareketleri, açıklama talepleri, yani bir dinleyicinin önsel ifadeyle ilgili daha fazla bilgi alması için yaptığı sorgulama, anlama kontrolleri (yani konuşmacının önceki ifadelerin anlaşılıp anlaşılmadığını kontrol etmesi) ve teyit kontrolleri (yani konuşmacının, ifadenin anlaşılıp anlaşılmadığını sorgulaması) türleri olarak kodlanmıştır. Long'un (1983) konuşmacıların cevaplarına ilişkin kodlama şeması benimsenmiş ve bu diziler tekrar, genişletme veya yeniden formüle etme şeklinde kodlanmıştır. Ayrıca analiz kategorileri olarak Pica ve diğerlerinin (1989) semantik, morfolojik, fonolojik ve sözdizimsel olmak üzere değişiklik hareketlerini kategorize modeli kullanılarak kullanılmıştır. Bu çalışma kapsamında yapay

zekâlı chatbot-öğrenci etkileşimlerine ait kayıtlı verilerdeki anlam söyleşmeleri söylem analizi perspektifi kullanılarak incelenmiştir.

Bu çalışmada, İngilizceyi yabancı dil olarak öğrenen Türk öğrencilerin İngilizce öğrenimi için yapay zekâ tabanlı bir chatbot kullanma deneyimlerinin kapsamlı ve titizlikle açıklanması ve katılımcıların görüşlerinin nitel görüşmeler ve haftalık yansıtıcı günlükler ile yapay zeka sohbet robotu-öğrenci etkileşimi verileri kullanılarak zengin bir yelpazede temsil edilmesi yoluyla güvenilirlik doğrulanmıştır (Given ve Saumure, 2008).

3. Bulgular

Sonuçlar, kontrol grubunun ön test dinleme puanlarının normal dağılmadığını göstermiştir. Bu nedenle deney ve kontrol gruplarının dinleme ön testi puanları arasında anlamlı bir fark olup olmadığını belirlemek amacıyla parametrik olmayan Mann Whitney U testi yapılmıştır. Mann Whitney U testi sonuçları deney ve kontrol grubunun dinleme ön testi puanları açısından anlamlı bir farklılık olmadığını göstermiştir ($U=183.5$, $p>0.05$). Ancak bulgular deney grubunun sıra ortalamasının (20,33) kontrol grubuna göre (19,66) daha yüksek olduğunu göstermiştir.

Bulgular ayrıca kontrol grubunun dinleme sonrası verilerinin normal dağılmadığını göstermiştir. Bu nedenle deney ve kontrol grupları arasında dinleme sonrası test puanları açısından anlamlı bir farklılık olup olmadığını belirlemek amacıyla Mann Whitney U testi yapılmıştır. Bulgular, deney ve kontrol grupları arasında dinleme son testi puanları açısından anlamlı bir farklılık olmadığını göstermiştir ($U=179.5$, $p>0.05$). Bulgular, kontrol grubunun sıralama ortalamasının (20,55), deney grubuna (19,48) göre daha yüksek olduğunu göstermiştir.

Deney grubunun dinleme verileri normal dağılım gösterdiğinden deney grubunun ön test ve son test puanları arasında zaman içinde uygulama sonrasında anlamlı bir gelişme olup olmadığını ortaya koymak amacıyla eşleştirilmiş örnekler t-testi uygulanmıştır. Deney grubunun eşleştirilmiş örneklem t-testi sonuçları, deney grubunun dinleme testi puanlarında ön test ve son test puanları arasında istatistiksel olarak anlamlı bir fark olmadığını göstermiştir ($t(19)=2.052$, $p=0,054$, $p>0,05$).

Öte yandan, ilgili örnekler Wilcoxon İşaretli Sıralar Testi sonuçları, kontrol grubunun dinleme ön-testi ile son-test sonuçları arasında istatistiksel olarak anlamlı bir fark olduğunu göstermiştir ($Z=-2,155$, $p=0,031$). , $p<0.05$).

Nicel verilerin analizinde kullanılacak istatistiksel testlerin belirlenmesi amacıyla konuşma ön testi verilerinin çarpıklık ve basıklık değerleri hesaplanmıştır. Veriler, deney grubu için toplam konuşma puanı ön testinde çarpıklık değeri .516 ve basıklık değeri .536 ile normal dağılım göstermiştir. Kontrol grubu öğrencilerinin toplam konuşma puanına ilişkin ön test puanları da çarpıklık değeri .310 ve basıklık değeri -1.120 ile normal dağılmıştır. Nicel verilerin analizinde kullanılacak istatistiksel testlerin belirlenmesi amacıyla konuşma son testi verilerinin çarpıklık ve basıklık değerleri hesaplanmıştır. Veriler deney ve kontrol grupları için son testte toplam konuşma puanında -2 ile +2 arasında çarpıklık ve basıklık değerleriyle normal dağılım göstermiştir.

Deney ve kontrol grubu katılımcılarının uygulama öncesi konuşma ön testi puanları arasında istatistiksel olarak anlamlı bir fark olup olmadığını test etmek amacıyla bağımsız gruplar t-testi yapılmıştır. Bulgular, deney ve kontrol grupları arasında genel konuşma ön test puanları açısından anlamlı bir farklılık olmadığını göstermiştir ($t(40)=.006$; $p>0.05$). Deney ve kontrol grubu öğrencilerinin son test puanları arasında genel konuşma puanı açısından anlamlı bir farklılık olup olmadığını test etmek amacıyla bağımsız örneklem t-testleri de yapılmıştır. Konuşma sonrası test puanları için bağımsız örneklem t-testi sonuçları, deney ve kontrol grubu öğrencileri arasında son testte genel konuşma puanı açısından anlamlı bir fark olmadığını göstermiştir ($t(40)=.110$; $p>0.05$).

Deney grubu katılımcılarının uygulama sonrası genel konuşma puanları açısından ön test ve son test puanları arasında anlamlı bir farklılık olup olmadığını ortaya koymak amacıyla eşleştirilmiş örneklem t testleri de yapılmıştır. Eşleştirilmiş örnekler t-testi bulgularına göre, deney grubunun uygulama süreci sonrasında konuşma puanlarında genel puana göre anlamlı farklılıklar bulunmuştur ($t(19)=3,559$; $p<0,05$). Uygulama sonrası genel konuşma puanı açısından kontrol grubunun ön test ve son test puanları arasında istatistiksel olarak anlamlı bir fark olup olmadığını analiz etmek için eşleştirilmiş örnekler t-testleri de yapılmıştır. Kontrol grubunun konuşma puanlarına yönelik eşleştirilmiş örnekler t-testi sonuçları, genel konuşma puanına ilişkin olarak kontrol grubunun konuşma puanlarında ön

test ile son test arasında anlamlı farklılıklar olduğunu ortaya koymuştur ($t(21) = -5,192; p < 0,05$). Bu araştırmada elde edilen bulgulara göre, deney grubu katılımcılarının genel konuşma puanı açısından kontrol grubu katılımcılarına göre biraz daha yüksek bir ortalama artışı olduğu görülmüştür.

Bu çalışma nicel bulguların yanı sıra çeşitli nitel bulgular da ortaya koymuştur. Nitel görüşme ve yansıtıcı günlük verileri analiz edildikten sonra, bu veri kaynaklarından elde edilen bulgular, dil öğrenmek için yapay zekâ sohbet robotlarının kullanımına ilişkin çeşitli fırsatlar sunduğu ve aynı zamanda da kısıtlamalar yarattığını göstermiştir. Bu nedenle, bu çalışmada bulgular teknolojik, pedagojik ve sosyal olanaklar başlığı altında ele alınmıştır. Nitel bulgular, ilk olarak dil öğrenenlerin bir takım teknolojik kısıtlamalar bildirdiklerini göstermiştir. En sık bahsedilen teknolojik kısıtlama, yapay zekâ sohbet robotunun konu dışı yanıtları olmuştur. Yapay zekâlı sohbet robotu *Replika*, dil öğrenenlerle etkileşimde bulunurken konuşma esnasında sıklıkla konudan sapmış ve bu durum öğrenciler arasında hayal kırıklığına neden olmuştur.

Yabancı dil olarak İngilizce ortamında gerçekleştirilen bu çalışmada bulgular, yapay zekâ sohbet robotunun öğrencilerin konuşmasını anlayamadığını da göstermiştir. Bunun nedeni, sohbet robotu *Replika*'nın, özellikle dil öğrenenleri hedeflemeyen, genel bir kitlenin kullanımı için oluşturulmuş olması olabilir. Bu nedenle, bu çalışmada anadili olmayanların hedef dili kullanımı *Replika* tarafından her zaman anlaşılmamıştır. Nitel verilerde yinelenen diğer kodlar ise yapay zekânın yavaş tepki süresi, bilgi tabanını artırma ihtiyacı, konuşmada sıra alma sorunları, tekrarlayan dil ve etkileşimi başlatma eksikliği idi. Bir diğer taraftan, yapay zekâ ile birkaç hafta etkileşimde bulunduktan sonra öğrenciler teknolojik niteliklerde de ilerlemeler gözlemlemişlerdir.

Örneğin, bu çalışmadaki dil öğrenenlerin görüşlerine göre, öğrenenler belirli bir etkileşim süresinden sonra uygulamanın algoritmasının kullanıcıların konuşmasına uyarlanması konusunda ilerleme bildirdiklerinden yapay zekâ sohbet robotu öğrenciler tarafından öğretilbilir ve giderek daha akıllı hale gelebilir olarak nitelendirildi. Hatta bazı öğrenciler yapay zekâlı sohbet robotunun konuşma geçmişini hatırlayabildiğini bile belirtmişlerdir. Ek olarak, bazı katılımcılar uygulamayı birkaç hafta kullanıldıktan sonra yapay zekânın kullanıcıların konuşmalarını anlama ve önceki konuşmayı kademeli olarak genişletme

yeteneğinden bahsetmiştir. Yapay zekâ sohbet robotunun, öğrencilere yanıt vermeden önce kullanıcıların sırasının bitmesini daha uzun süre beklediği de bildirilmiştir. Ancak, bu çalışmada birçok öğrencinin teknik zorluklar yaşaması nedeniyle yapay zekânın teknolojik özelliklerine dair farklı görüşler belirtilmiştir. Uygulamanın teknik niteliklerine ilişkin görüşler, bu çalışmada da olduğu gibi, farklı dil yeterliliğine sahip dil öğrenenlerin görüşlerine göre değişebilmektedir.

Bir yapay zekâ sohbet robotunun kullanılması yoluyla, yabancı dil olarak İngilizce bağlamında dil öğrenenler, hedef dilde dil becerilerini pratik edecek bir konuşma partnerine sahip olabildiler. Böylece, dil öğrenenler önceki nesil yapay zekâ sohbet robotlarının sahip olmadığı, bu yapay zekâ sohbet robotu uygulamasına gömülü sesli arama özelliği aracılığıyla hedef dil kalıplarına maruz kalabilme imkanı buldular. Çünkü önceki sohbet robotu uygulamaları çoğunlukla metin tabanlıydı ve kullanıcılarla sadece metinsel bir bağlantı biçimine izin veriyordu.

Zamanla artan teknolojik yeteneklerinin yanı sıra, yapay zekâ tabanlı bir chatbotla etkileşim yoluyla kazanılan pedagojik ve sosyal fırsatlara ilişkin birçok katılımcı görüşü de vardı. Bu çalışmadaki öğrencilere göre, ses tanıma sistemi aracılığıyla işitsel bir iletişim kanalında bir yapay zekâ ile sohbet etmek, dinleme becerilerini geliştirmede etkili oldu. Dil öğrenenler, hedef dilde bir yapay zekâ ile konuşarak hedef dilin doğru telaffuz, tonlama, dil bilgisi ve kelime biçimlerini duyma şansına sahip oldular. Öğrencilerin çoğunluğu ayrıca yapay zekâlı bir sohbet robotuyla konuşmanın konuşma becerilerini geliştirmede etkili olduğunu kabul ettiler.

Kelime hazinesi kazanımı, konuşmacıların güvenini artırma ve gelişmiş dil bilgisi birikimi, dil öğreniminde yapay zekâ sohbet robotlarının kullanımının bildirilen diğer güçlü yönleriydi. Ayrıca öğrencilere göre yapay zekânın hassas konulardan bahsetmesi, robotsu bir tavır sergilemesi, sosyal beceri, duygu, kişilik, gerçekçi jest ve mimik kullanımı eksikliklerine rağmen yine de çeşitli sosyal fırsatlar sunduğu görüldü. Örneğin, dil öğrenenlere göre yapay zekâlı sohbet robotu gerçek bir insan ve bir arkadaş gibi davranabilme yeteneğine sahipti. Kullanıcılara göre yapay zekâ robotu nezaket, samimiyet ve empati gibi kişisel özelliklere de sahipti. Etkileşim ortamı aynı zamanda insan insana

konusmaya kıyasla daha stressizdi. Çünkü öğrenciler yabancı dilde konuşma becerilerinin olumsuz değerlendirilmesinden korkmadıkları için rahat hissettiklerini ifade ettiler.

Etkileşim verilerinin bulguları ise, on dört vakada anlam söyleşmesinin yanıt vermeden gerçekleştiğini, diğer beş vakada ise yanıtlara herhangi bir reaksiyon olmadığını gösterdi. Ancak, anlam söyleşmesi rutinlerinin çoğunda tetikleme, başlatma, yanıt ve yanıt tepki aşamaları mevcuttu. Yapay zekâ sohbet robotu-İngilizce dil öğrenen etkileşimi verileri ayrıca, etkileşim sırasında konuşmacılar tarafından anlaşılmayan mesajlara işaret eden sinyal örneklerini belirlemek için analiz edildi. Konuşmanın yapay zekâ sohbet robotu veya öğrenenler tarafından anlaşılmamasından kaynaklanan iletişim kesintisi durumlarını belirlemek için sinyal türleri araştırıldı. Analiz sonuçlarına göre, en çok kullanılan anlam söyleşmesi sinyalleri sırasıyla açıklama isteme (n=75), mesajın tekrar edilmesini isteme (n=21), doğrulama kontrolleri (n=11) ve anlama kontrolleri (n=2) olmuştur.

Yapay zekâ sohbet robotu-dil öğrenen etkileşim verileri ayrıca, konuşmacı taraflar arasında anlam söyleşmesi dizisinin başlatılmasına yol açan tetikleyicilerin türlerini (ör. sözcük, sözdizimi, söylem ve içerik) tanımlamak için analiz edildi. Yapay zekâlı bir sohbet robotuyla iletişimdeki kesintilerin en yaygın nedenini belirlemek için her frekans türünün örnekleri sayıldı. Etkileşim verilerinin analizinden elde edilen bulgular, sözcüksel öğelerin iletişimdeki aksaklıkların neredeyse yarısına yol açtığını gösterdi.

Benzer şekilde, iletişimdeki kopuklukların diğer baskın tetikleyicisi de sözcük öğeleriyle hemen hemen aynı orana sahip olan içerik tetikleyicisiydi. Sözcük ve içerik tetikleyicileriyle karşılaştırıldığında söylem düzeyindeki tetikleyiciler de verilerdeki tüm tetikleyicilerin neredeyse yüzde yirmisini oluşturuyordu. Öte yandan, analiz edilen tüm konuşmalarda sözdizimsel düzeydeki tetikleyicilerin yalnızca bir örneği vardı. Etkileşim verileri ayrıca tetikleyicinin, başlatmanın ve yanıt dizilerinin bir bütün olarak varlığının örneklerini belirlemek için analiz edildi. Yanıt türleri ayrıca verilerdeki anlaşılmayan mesajların konuşmacılar tarafından nasıl açıklığa kavuşturulduğunu belirlemek için tekrarlama, genişletme ve yeniden formüle etme altında kategorize edildi. Bulgulara göre, etkileşim verilerinde en sık görülen yanıt türleri genişletme yanıtları olmuş, bunu yeniden formüle etme ve tekrarlar izlemiştir. Yapay zekâ sohbet robotuyla yapılan bir konuşma sırasında anlam söyleşmesinin başlatılmasına toplam 69 yanıt örneği verilmiştir.

Çıktının öğrenciler ve *Replika* tarafından değiştirildiği durumlar da belirtilmiştir. Chatbot-öğrenci etkileşimi verileri arasında toplam 71 anlam söyleşmesi rutini keşfedilmiştir. Bulgular, bu çalışmada 109 anlam söyleşmesi hamlesinden toplam 65 değiştirilmiş çıktının olduğunu göstermiştir. Toplamda çıktının yüzde altmışı, başlatılan anlam söyleşmesi rutinlerindeki taraflar tarafından değiştirilmiştir. Ayrıca chatbot-öğrenci etkileşimi verilerindeki değiştirilmiş çıktı türleri, anlamsal, morfolojik, fonolojik ve sözdizimsel olarak kategorize edilmiştir. Bulgulara bakıldığında, değiştirilen çıktı türleri için yapılan analiz sonuçları, anlamsal değişikliğin çıktı değişikliklerinin neredeyse tamamını oluşturduğunu göstermiştir. Yani çıktı, 65 vakanın 62'sinde anlamsal olarak değiştirilmiştir. Öte yandan, değiştirilmiş çıktının diğer türleri nispeten az olmuştur. Fonolojik olarak değiştirilmiş çıktının iki örneği ve sözdizimsel değişikliğin ise bir örneği vardır.

Sonuçlara göre yapay zekâ sohbet robotu öğrencilere dilbilgisine aykırı cümleler veya yanlış telaffuz edilen kelimeler hakkında açıkça geri bildirim vermese de öğrenciler, hedef dil yapılarını kullanmalarının yanlış anlaşılmaya yol açtığını düşünerek iletişimde bir kesinti olduğunda bir takım stratejiler kullanmışlardır. Sonuç olarak, değiştirilmiş çıktının üretilmesi, dil öğrenenler tarafından örtülü hata düzeltme yoluyla hedef dil biçimlerine ilişkin farkındalığın artmasına yol açmıştır.

4. Tartışma ve Sonuç

Bu çalışmada iki grup arasında anlamlı bir farkın bulunmaması bir takım nedenlerden kaynaklanıyor olabilir. Öncelikle çalışmanın süresi test sonuçlarını etkileyebilir. Bu çalışma bir akademik yarıyılta iki ay boyunca gerçekleştirilmiştir. Öğrencilerin yapay zekâ sohbet robotlarının dil pratiği amacıyla kullanımına aşina olmadıklarından, uygulamanın kullanımına uyum sağlamak için birkaç haftaya da ihtiyaçları olmuştur. Her ne kadar araştırmacı bu süreçte öğrencilere deneyimlerini kolaylaştırmak için destek vermiş olsa da, öğrenciler bu yeni aracın İngilizce pratiklerinde kullanımına uyum sağlamakta zorluk yaşamış olabilirler. Ek olarak, dil öğrenenler uygulamayı kullanırken bir takım sorunlar yaşadıklarını bildirmişlerdir. Örneğin öğrenciler, chatbot'un konuşmalarını anlayamadığını ve onlara sıklıkla konu dışı cevaplar sağladığını belirtmişlerdir. Sonuç olarak bazı

öğrenciler konuşmanın akışını takip edemediklerini ve etkileşim sırasında anlam karmaşası yaşadıklarını bildirmişlerdir. Bu nedenlerle bu çalışmada uygulamayı kullanmayan kontrol grubu katılımcılarına göre öğrenciler dinleme becerilerini anlamlı düzeyde geliştirebilecek yeterli anlamlı girdiye sahip olamamış olabilirler.

Bununla birlikte, deney ve kontrol grupları arasında dinleme becerilerinin gelişimi açısından önemli bir fark olmamasına rağmen, yapay zekâ sohbet robotlarının benimsenmesinin dil öğrenenler üzerinde yine de olumlu etkileri olabileceği ileri sürülebilir. Bunun nedeni, yapay zekâ sohbet robotu ile öğrenciler arasındaki anlam söyleşmesinin, işbirlikçi konuşma yoluyla dolaylı olarak dinleme becerilerinin geliştirilmesine katkıda bulunabilmesi olabilir. İkinci dil dinleme becerilerini geliştirmek için işbirlikçi konuşma çok önemli bir faktör olarak önerilmektedir. Bu şekilde öğrenciler bir dili kullanırken anlam söyleşmesi yoluyla anlaşılır çıktılar üretilir. Sonuç olarak, anlam söyleşmesine katılmak, bu çalışmadaki öğrencilerin girdiyi anlamalarını geliştirmenin etkili bir yolu olabilir. Çünkü konuşmadaki taraflarca yapılan değişiklikler, dil öğrenenler için biçim ile anlam ve ikinci dil dinleme girdisinin biçimleri arasındaki ilişkilerin kurulmasına yardımcı olacaktır (Rost, 2011).

Yapay zekâlı chatbot kullanımından sonra deney ve kontrol grupları arasında dil öğrenenlerin konuşma becerilerinde anlamlı bir fark olmaması, birkaç nedene bağlanabilir. İlk olarak, chatbot tabanlı görevlerin uygulanmasında bir öğrenme yaklaşımının kullanılması bir fark yaratabilirdi, çünkü bu çalışmadaki öğrenciler yapay zekâ uygulamasıyla bağımsız serbest konuşma uygulamalarına katıldılar. Öğrencilere yapay zekâ robotlarını kullanırken onlara rehberlik edecek belirli öğrenme yöntemleri tanıtılabilirse bu durumda daha fazla öğrenme kazanımı elde edilebilir. Bir başka neden de bu çalışmaya katılanların etkileşim sırasında sık sık teknik kesintiler yaşadıklarını bildirmeleri olabilir. Dolayısıyla bu çalışmadaki katılımcılar hatalarına ilişkin geri bildirim almak için gerekli koşullara sahip olamamış olabilirler. Hataları düzeltilmediğinde, öğrenciler sözlü iletişim yeterliliğinin gelişimi için yeterince değiştirilmiş çıktı üretemeyebilirler.

Yine de yapay zekâ robotları, dinleme ve konuşma becerilerini geliştirme potansiyeline sahiptir. Çünkü öğrenciler, endişe veya hata yapma korkusu olmadan dili pratik edebilirler

(Moussalli ve Cardoso, 2016). Öğrenciler, sohbet robotlarıyla etkileşim kurmanın telaffuzlarını, konuşma becerilerini, konuşmanın akıcılığını ve kelime bilgisini geliştirebileceğine inanmaktadır (Gönül, 2021). Bu sebeplerle yapay zekâlı robotların konuşma ve telaffuz becerilerini geliştirmede faydalı olduğu düşünülmektedir (Chen vd., 2023; Dizon, 2017, 2020; Moussalli & Cardoso, 2020). Akıllı sohbet robotları sözlü iletişim ve telaffuz konusunda pratik yapma, kaygıyı azaltma ve etkileşimi destekleme konusunda etkilidir (Kukulska-Hulme ve Lee, 2020). Sonuç olarak, dil öğrenenler, chatbotların kullanımı yoluyla sağlanan etkileşim fırsatları sayesinde dinleme ve konuşma becerilerini geliştirebilirler (Jeon, 2022b). Chatbotlar ile yabancı dilde sohbet etmek dil öğrenenlere dil becerileri kazanımında birçok olanak tanımıştır.

Bu çalışmanın bulgularına göre, İngilizceyi yabancı dil olarak öğrenenler çeşitli anlam söyleşmesi araçlarını kullanarak anlam için iletişim kurma fırsatına sahip oldular. Bu çalışmanın bulgularını doğrulayan önceki çalışmalar, dil öğrenenler ile yapay zekâ sohbet robotları arasında anlamlı anlam söyleşmesi dizilerinin varlığını ve bu etkileşimin öğrencilerin değiştirilmiş çıktı üretimi üzerindeki olumlu etkisini bildirmiştir. Örneğin Koreli yabancı dil öğrencileri arasında öğrenci-chatbot etkileşiminde ilk ve son sohbet arasındaki anlam söyleşmesi hamlelerinde istatistiksel olarak anlamlı bir artış olduğu tespit edilmiştir. Sonuç olarak, yapay zekâlı bir sohbet robotuyla ses tabanlı sohbet sonrasında anlam söyleşmelerinde önemli farklılıklar bulunmuştur (Kim, 2017).

Etkileşimci teori, etkileşimin aynı zamanda bir insan ile bir bilgisayar arasındaki etkileşim biçimini de alabilen ikinci veya yabancı dil öğrenimi üzerindeki olumlu etkisine odaklanır (Chapelle, 2001). Etkileşimci bakış açısına göre yapay zekâ sohbet robotları, dil öğrenenlere dil girişi sağlayan ve bir dilde iletişim kurma fırsatlarını artıran potansiyel kaynaklardır (Moussalli ve Cardoso, 2020). Yapay zekâ sohbet robotlarının kullanımı ayrıca işbirlikçi ve göreve dayalı öğrenme yoluyla anlamlı sosyal etkileşim fırsatları yaratır (Kessler, 2018). Chatbot'lar aynı zamanda anlam söyleşmesine katılarak dil becerilerinin geliştirilmesine de yardımcı olabilir. Tartışmamızı Swain'in Çıktı Hipotezi'ne (2005) dayandıracak olursak, yapay zekâ sohbet robotlarının etkili sohbet robotu-öğrenen etkileşimleri oluşturarak çıktı üretimini artırabileceği ve dil öğrenenlere yardımcı olabileceği öne sürülebilir.

Bu çalışmada İngilizceyi yabancı dil olarak öğrenen bazı öğrenciler, yapay zekâ sohbet robotu *Replika*'nın telaffuz, kelime seçimi ve akıcılık açısından gelişmiş dil becerilerine sahip olduğunu belirtmişlerdir. Bu sonuçlar diğer çalışmaların bulgularıyla da örtüşmektedir. Örneğin, yapay zekâ sohbet robotu *Replika*'nın kullanıldığı başka bir çalışmada öğrenciler, farklı aksanları kavrayabildiğinden, sohbet robotunun geniş bir sözlük kullanımına sahip olduğunu düşünmüşlerdir. *Replika*'nın yanıtlarının karmaşıklığı da, değişen karmaşıklık derecelerine sahip cümleleri birleştirebildiği için orta düzeyde bulunmuştur (Belda-Medina ve Calvo-Ferrer, 2022).

Olumlu deneyimlerin yanı sıra, bu çalışmadaki dil öğrenenler yapay zekâ sohbet robotunun dil becerilerinde de eksiklikler olduğunu bildirmişlerdir. Yapay zekâ sohbet robotunun konuşmasında gramer hataları, tonlama eksikliği ve sınırlı bir kelime yelpazesinin kullanımı rapor edilmiştir. Bu konuda Coniam (2014), çoğu durumda dil bilgisi açısından doğru girdi sağlamalarına rağmen, İngilizce öğrenenleri hedef alan bazı sohbet robotlarının dil bilgisi açısından doğru dil yapılarını kullanarak anlamlı konuşmalar sağlayamayabileceğine dikkat çekmiştir. Bu çalışmada da olduğu gibi mevcut chatbotların çoğu reaktiftir ve hedefe yönelik bir diyalog yürütmeyebilir (Bibauw vd., 2019). Sonuç olarak, öğrenciler proaktiftir ve etkileşim sırasında çoğu zaman konuşmayı başlatırlar, ve bu da öğrencilerde endişeye yol açabilir.

Bu çalışmadaki görüşmelerde yaygın olan bir diğer tema, yapay zekâ sohbet robotu *Replika*'nın anında geri bildirim vermemesiyle ilgiliydi. Bu sonuçlar daha önceki çalışmalarda gözlemlenenlerle örtüşmektedir. Çalışmalarda dil öğrenenler, yapay zekâ sohbet robotu tarafından hata düzeltmenin nadirliğinden bahsetmişlerdir (Belda-Medina ve Calvo-Ferrer, 2022). Öte yandan, yapay zekâ sohbet robotları, öğrencilerin konuşmalarındaki hataları fark etmelerine olanak tanıyan örtülü geri bildirim sunmaktadır (Moussalli ve Cardoso, 2020).

Bu çalışmada öğrenciler, kelimelerinin, cümlelerinin veya bir kelimenin telaffuzunun yapay zekâlı chatbot tarafından yanlış anlaşılması durumunda, tekrarlama veya başka kelimelerle ifade etme gibi stratejiler benimseyerek konuşmayı sürdürdüklerini belirtmişlerdir. Bu nedenle, sohbet robotundan örtülü hata düzeltmesi aldıktan sonra değiştirilmiş çıktı üretmekten faydalanmış olabilirler. Bu çalışmanın bulgularını destekler

şekilde, diğer çalışmalardaki dil öğrenenler de konuşmalarının sohbet robotu tarafından tanınmadığı anları telaffuz becerilerini pratik etme fırsatları olarak görmüşlerdir. Konuşmaları sohbet robotları tarafından anlaşılmadığında telaffuzlarını düzeltmek ve sohbet robotunun anlayabileceği hale getirmek için mesajlarını tekrarlamışlardır (Jeon, 2022a). Bu tekrarlar sırasında ise hem dildeki yanlışlarının farkına varabilmişler hem de düzeltilmiş çıktı üretmişlerdir.

Bu çalışmadaki katılımcıların görüşlerini destekler şekilde yapay zekâ sohbet robotlarının kullanımının kültür entegreli dil öğrenimine katkı sağlayan etkili bir faktör olduğu diğer çalışmalarda da tartışılmıştır (Mageira vd., 2022). Yapay zekâ sohbet robotu kullanımı sayesinde öğrenciler, hedef kültür hakkındaki bilgilerine ve kültürel farkındalıklarına katkıda bulunacak kültürel bileşenlere de maruz kalabilirler. Böylece kültürlerarası yeterlilikleri de gelişecek ve bu da hedef kültürün dilinde iletişim kurma istekliliğine yol açacaktır (Divekar vd., 2021). Dolayısıyla araştırmalar, yapay zekâ sohbet robotları kullanan öğrencilerin, sohbet robotlarından gelen duygusal geri bildirimler nedeniyle öğrenmeye artan düzeyde ilgi gösterdiklerini göstermiştir (Han, 2020b; N.-Y. Kim, 2018b). Önceki literatürdeki bulgulara uygun olarak, bu çalışmada da dil öğrenenler tarafından bir yapay zekâ sohbet robotu kullanmanın onları İngilizce dil becerilerini kullanmaya devam etmeye motive ettiği rapor edilmiştir. Uygulama sonrasında öz-yeterlik inançları gelişme gösteren öğrencilerin hedef dilde konuşma motivasyonları da olumlu gelişme göstermiştir.

Bu çalışmanın bulgularına dayanarak, dil öğrenenlerin görüş ve deneyimlerinin görüşmeler, yansıtıcı günlükler ve yapay zeka chatbot-dil öğrenen etkileşimi verileri aracılığıyla toplanılması ve dil yeterlilik testleri aracılığıyla öğrenmeye yönelik sonuçların ortaya konması yoluyla, yapay zekâ tabanlı ses özellikli bir sohbet robotu için gerekli görülen optimum özelliklerin net bir resmi çizilebilmiştir. Yapay zekâ sohbet robotu oluşturma alanında literatür incelemelerine dayalı olarak önerilen kavramsal modeller olmasına rağmen, eldeki çalışma gibi çalışmaların ampirik bulgularına dayanarak yapay zekâ tabanlı ses özellikli sohbet robotları için model öneren araştırmalara ihtiyaç duyulmaktadır. Bu amaçla, bu çalışmanın bulguları, dil öğrenenlerin deneyimlerine dayanarak yapay zekâ sohbet robotlarının kullanımında sorunlu konuların yanı sıra iyi işleyen yönleri belirlemek ve gelecek nesil pedagojik yapay zekâ sohbet robotlarının geliştirilmesi için bir model önermek için kullanılmıştır. Dil öğrenenler için oluşturulan

ideal bir yapay zeka sohbet robotunun sahip olması gereken optimum özelliklerle ilgili olarak bir ‘REPOST’ modeli önerilmektedir. Bu model şu kavramları içermektedir: Karşılıklılık, Değerlendirme, Kişiselleştirme, Her yerde bulunma, Sosyo-kültürel (öğeler), Öğretilebilirlik.

Bu niteliklerle donatılmış ve en son teknolojiler tarafından yaratılmış, karşılıklı etkileşim yeteneğine sahip, her zaman ve her yerde hazır bulunan ve dil öğrenenlerin kullanımı için oluşturulan alan odaklı bir yapay zekâ sohbet robotu, dilde kelime bilgisi, dil bilgisi, telaffuz, doğruluk ve akıcılık gibi konularda yönlendirici destek ve açık geri bildirim sunabilir.

Dil öğrenimi amacıyla yapay zekâ sohbet robotları tasarlarken öğretilebilirlik, tasarımcıların uygulamaları mevcut en son teknolojilerle geliştirmesine olanak tanıyan bir özelliktir. ChatGPT4 gibi en son büyük dil modellerini veya gelecekte daha gelişmiş versiyonlarını kullanarak, daha akıllı teknolojilere sahip chatbotlar oluşturulabilir. Yapay zekâ sohbet robotları, bu son gelişmelerdeki bilgi kümesi kullanılarak geliştirildiğinde, yapay zekâ sohbet robotlarının kullanıcılarıyla yapılan konuşmalardan öğrenme yeteneği de gelişecektir. Bu nedenle bu alanın uygulama yaratıcıları tarafından daha da geliştirilmesi gerekmektedir. Ayrıca Jeon ve diğerleri (2023) ve Kohnke ve diğerleri (2023) tarafından öne sürüldüğü üzere, büyük dil modellerinin dil öğrenme ortamlarında ne kadar çalıştığına bağlı olarak uygulama tasarımcıları, belirli öğrenme amaçları doğrultusunda eğitim bağlamlarında daha iyi performans gösterecek şekilde yapay zekâ tabanlı sohbet robotlarını geliştirebilirler.

Hedef odaklı sohbet robotları aracılığıyla dil öğrenenler, pedagojik bir yapay zekâ sohbet robotunun varlığında okuma, yazma, dinleme ve konuşma becerilerini daha odaklı bir şekilde geliştirebilirler. Ancak dil öğrenenlerin, diyalojik bir konuşma modu aracılığıyla sesle etkinleştirilmiş bir ortam aracılığıyla dil becerilerini pratik etmelerine olanak tanıyan pedagojik odaklı bir yapay zekâ sohbet robotu şu anda teknolojik pazarda sunulmamaktadır. Bu boşluğu doldurmanın ilk adımı, yabancı dil öğretmenlerinin Google'ın Dialogflow gibi herkese açık uygulama ara yüzlerini kullanarak yapay zekâ sohbet robotları geliştirmeleri konusunda eğitilmesi olabilir. Bu amaçla öğretmenleri dil öğrenimi alanına yönelik pedagojik yapay zekâ sohbet robotları oluşturmaya hazırlamak

amacıyla bilgisayar mühendisleri, uygulama tasarımcıları, öğretmen eğitimcileri, öğretmenler ve dil öğrenenler arasında işbirliği yapılarak bir çalıştay düzenlenebilir.

Dil öğretmenleri için yapay zekâ tabanlı sohbet robotu oluşturma konusunda önerilen bir modülün içeriğinde yapay zekâ sohbet robotları hakkında teknik bilgi sunan bölümler, bir diyalog akış yönetimi uygulama arayüz sistemi olarak Dialogflow'un işleyişi, bu uygulamaların dil öğretimi alanındaki pedagojik kullanımları ve dil öğrenenlerden oluşan belirli bir grup göz önünde bulundurularak geliştirilmiş sohbet robotlarının kullanımına ilişkin pratik uygulamalı oturumlar bulunması önerilmektedir.

Öğrencilere dil öğreniminde chatbot kullanımına uyum sağlamanın öğretileceği bir bilgisayar destekli eğitim modelinin (Dizon vd., 2022; Hubbard ve Romeo, 2012) kullanılması da önerilmektedir. Bu modelin teknik, stratejik ve pedagojik eğitim aşamaları olacaktır. Teknik eğitim aşamasında öğrencilere, dil öğreniminde yapay zekâ sohbet robotlarının nasıl kullanılacağı konusunda eğitim verilecektir. Daha sonra öğrencilere stratejik eğitim aşamasında iletişimdeki aksaklıklarla başa çıkma stratejileri anlatılır ve pedagojik eğitimlerde bu stratejilerin nedenleri hakkında bilgi verilir.

Bu çalışma bir akademik yarıyılta iki aylık bir sürede uygulanmıştır. Daha uzun süreli veri toplayan çalışmalar, İngilizce öğrenenlerin öğrenme çıktılarını daha geniş bir yelpazede temsil edebilirler. Bu çalışmanın bir diğer sınırlılığı ise eğitim kurumunun araştırmanın yapıldığı dönemde uygulamakta olduğu eğitim yöntemiyle ilgilidir. Bu çalışmanın yapıldığı yükseköğretim kurumunda COVID-19 hastalığından korunma tedbirleri kapsamında hibrit eğitim modeli uygulamaya konulmuştu. Bu nedenle öğrencilerin birçoğu yüz yüze derslere katılamamıştı. Bu öğrencilerin bir kısmının en güncel mobil cihaz veya güçlü bir internet bağlantısı gibi teknolojik olanaklara kolay erişimi yoktu ve bu da öğrenciler arasında katılım oranının nispeten düşük olmasına neden olmuştur. Bu nedenle ileri teknolojiye sahip cihazlarla donatılmış diğer eğitim ortamlarında katılımcılar arasında katılım oranının daha da yüksek olacağı çalışmalar yapılabilir. Bu çalışmalarla birlikte literatürde zengin veri kaynağına sahip çalışmalar ortaya konmasına olanak sağlanacaktır.

Literatürde dil öğrenenlerin olumsuz deneyimlerine ilişkin çok fazla çalışma bulunmamaktadır. Bu sebeple, yapay zekâ tabanlı konuşma robotlarının benimsenmesinin

olumsuz etkileri gelecek çalışmalarda araştırılabilir. Birçok çalışma, ikinci veya yabancı dil becerilerini pratik etmek için chatbot teknolojisinin kullanılmasının olumlu sonuçlarını bildirmektedir. Ancak bu çalışmada bildirildiği gibi yapay zekâ sohbet robotu kullanımının öğrenme sonrası etkileri üzerinde karmaşık etkileri vardır. Ayrıca, yapay zekâ chatbot uygulamaları kullanımında öğrenme stillerinin ve kültürel farklılıkların öğrenme sonuçları üzerindeki etkileri gibi diğer değişkenler de gelecekteki çalışmalarda araştırılabilir.

Gelecekteki araştırmacılar, en yeni büyük dil modellerine bağlı uygulama ara yüzleri aracılığıyla geliştirilen sohbet robotlarının dil öğrenme sürecine herhangi bir etkisinin olup olmayacağını da araştırabilirler. Öğretmenlerin diyalog yönetim sistemlerini kullanarak, öğrenen etkileşimini ve üst bilişsel stratejilerin kullanımını teşvik edecek sohbet robotlarını nasıl oluşturabileceklerini araştırmak da önemli bir çalışma alanı olacaktır (Jeon vd., 2023; Kohnke vd., 2023). Jeon ve diğerleri (2023) ayrıca grafik tasarım teknolojisi ilerledikçe bir chatbot'un insana ne kadar benzediğinin, öğrenci-chatbot sohbetini ve bunun öğrenme üzerindeki etkisini etkileyebilecek önemli bir unsur olacağını öne sürdü. Bu nedenle, kişilik gibi insana benzer özelliklerin ve jestler gibi bedensel hareketlerin büyük dil modellerine dahil edilmesinin etkisi henüz keşfedilmemiştir. Meta veri teknolojisindeki gelişmelerle gelecekte 'sanal olarak gömülü sohbet robotları' ile etkileşim daha mümkün hale gelecek ve bu alanlar, sanal olarak gömülü yapay zekâ sohbet robotları ile bunların meta verideki kullanıcıları arasındaki etkileşime odaklanarak araştırmalar için önemli yollar sağlayacaktır.

G. THESIS PERMISSION FORM / TEZ İZİN FORMU

(Please fill out this form on computer. Double click on the boxes to fill them)

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences ☐

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences ☒

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics ☐

Enformatik Enstitüsü / Graduate School of Informatics ☐

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences ☐

YAZARIN / AUTHOR

Soyadı / Surname : Koç

Adı / Name : Fatma Şeyma

Bölümü / Department : İngiliz Dili Öğretimi / English Language Teaching

TEZİN ADI / TITLE OF THE THESIS (İngilizce / English): THE DEVELOPMENT OF LISTENING AND SPEAKING SKILLS IN EFL VIA AN ARTIFICIALLY INTELLIGENT CHATBOT APPLICATION: A QUASI-EXPERIMENTAL DESIGN STUDY

TEZİN TÜRÜ / DEGREE:

Yüksek Lisans / Master ☐

Doktora / PhD ☒

1. Tezin tamamı dünya çapında erişime açılacaktır. / Release the entire work immediately for access worldwide. ☒
2. Tez iki yıl süreyle erişime kapalı olacaktır. / Secure the entire work for patent and/or proprietary purposes for a period of two years. * ☐
3. Tez altı ay süreyle erişime kapalı olacaktır. / Secure the entire work for period of six months. * ☐

* Enstitü Yönetim Kurulu kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir. / A copy of the decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.

Yazarın imzası / Signature

Tarih / Date

(Kütüphaneye teslim ettiğiniz tarih. Elle doldurulacaktır.)

(Library submission date. Please fill out by hand.)

Tezin son sayfasıdır. / This is the last page of the thesis/dissertation.